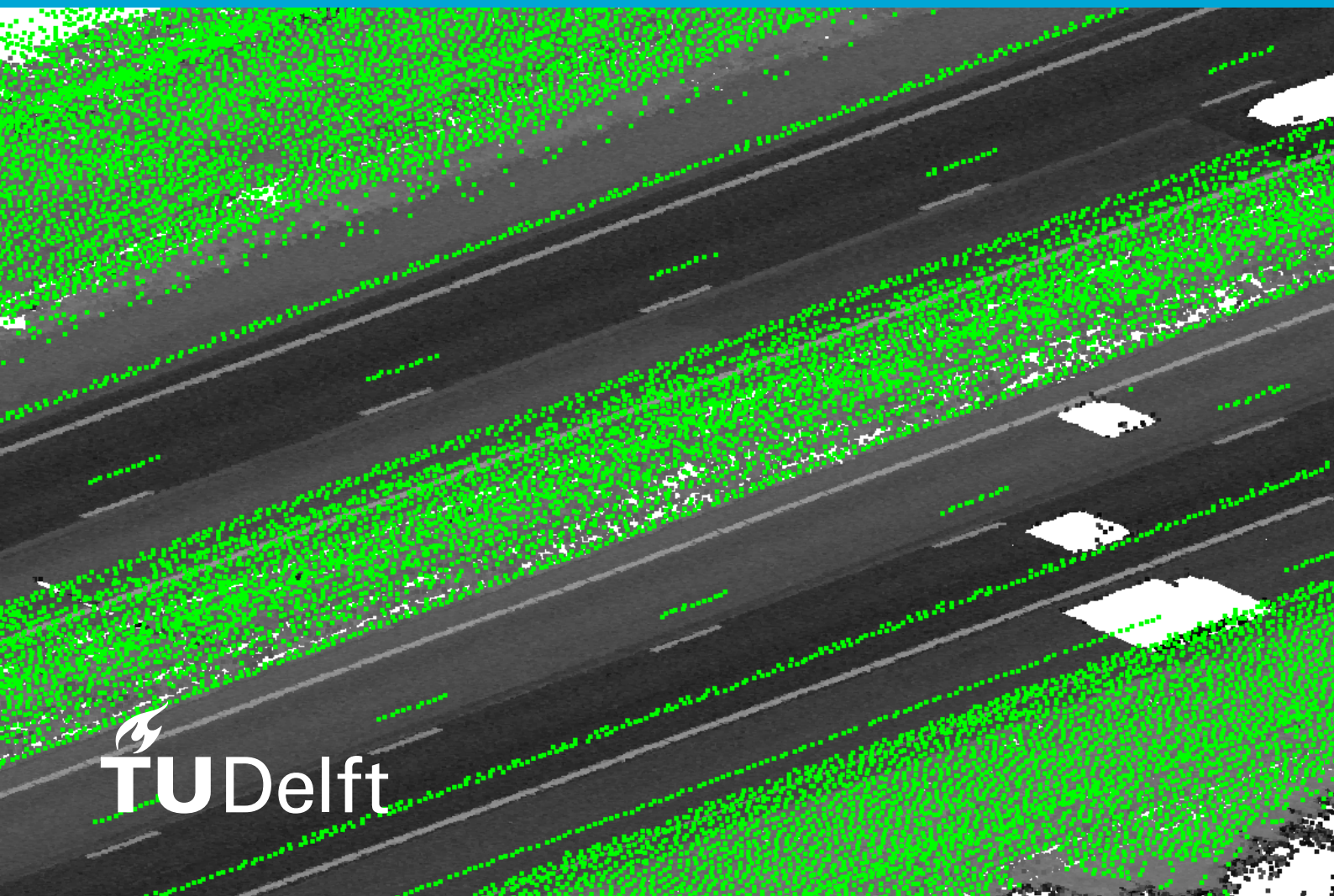
MSc thesis in Geomatics

Harmonisation of Heterogeneous Point Cloud Using Road Marking as Benchmark

Der Derian Auliyaa Bainus

2025



TU Delft

MSc thesis in Geomatics

Harmonisation of Heterogeneous Point Cloud Using Road Marking as Benchmark

Der Derian Auliyaa Bainus

June 2025

A thesis submitted to the Delft University of Technology in partial fulfilment of the requirements for the degree of Master of Science in Geomatics

The work in this thesis was supported by:

3D geoinformation group
Delft University of Technology

Rijkswaterstaat
Ministry of Infrastructure and Water Management

Supervisors:	ir. D.H. van der Heide
	Prof. dr. J.E. (Jantien) Stoter
Co-reader:	Shenglan Du

Abstract

Accurate alignment of heterogeneous LiDAR point clouds is important for producing high-quality elevation data. This process, known as harmonisation, corrects spatial discrepancies between overlapping datasets collected at different times or using different sensors. A key step in harmonisation is the use of reliable benchmarks, which are features that are clearly detectable in both datasets, to guide the alignment. While artificial targets have been traditionally used as such benchmarks, their deployment is costly and impractical for large-scale or uncoordinated surveys. Consequently, there is growing interest in extracting natural or man-made features directly from LiDAR data to serve as benchmarks.

Road markings have recently been proposed as an alternative benchmark due to their consistent visibility in LiDAR point clouds, particularly through intensity value. As part of the Integrale Hoogtevoorziening Nederland (IHN) initiative, road markings are being explored as benchmarks for national point cloud alignment. However, the performance of the road markings as a co-registration benchmark has not been well researched yet.

This research investigates how the accuracy of automatically extracted road markings affects the quality of point cloud alignment. It builds upon an adaptive extraction method that adjusts intensity thresholds to suit different datasets and applies geometric filtering to generate 3D line representations of road markings. These extracted features are used to align heterogeneous LiDAR point clouds, and their performance is evaluated against manually digitised road markings to assess how extraction quality influences alignment accuracy. To support this evaluation, a RANSAC-weighted centroid alignment approach is proposed, which uses the inlier count of each correspondence as a weight during transformation estimation, aiming to prioritise geometrically stable benchmarks in the alignment process.

The results show that alignment accuracy is influenced by the extraction quality and distribution of road markings. When fewer road markings are available, especially in datasets with larger time gaps due to environmental changes, the spatial distribution has a higher chance of becoming uneven, leading to transformation errors. These errors grow with distance from the benchmark area, increasing alignment inaccuracies across the dataset.

Acknowledgements

The completion of this thesis marks the end of my journey towards obtaining a Master's degree in Geomatics. This journey would not have been possible without the support of the Indonesia Endowment Fund for Education or Lembaga Pengelola Dana Pendidikan (LPDP), whose scholarship enabled me to pursue my studies at TU Delft. Along the way, several individuals and institutions have also played a significant role, from inspiring my desire to pursue a Master's degree to guiding me through to the completion of this thesis.

First and foremost, I would like to thank Daan van der Heide for giving me the opportunity to work on the topic of point cloud harmonisation and for providing valuable insights into the recent development of a method to extract co-registration benchmarks from point cloud data and its potential for improvement. These insights formed the foundation of the research and helped shape the development of ideas throughout the project. Through weekly discussions, I gained a lot of new knowledge, which helped me work on the project.

Furthermore, I would like to thank Jantien Stoter for providing valuable insights and guidance throughout the graduation process. The thoughtful feedback and broad perspective offered helped me to think more deeply about my work and understand the topic better. I would also like to thank Shenglan Du for taking the time to review my thesis report and provide insightful feedback.

I would also like to thank Deni Suwardhi, my undergraduate thesis supervisor at the Bandung Institute of Technology, for introducing me to the field of LiDAR point clouds and encouraging my interest in exploring it further through higher education. In addition, I am grateful to PT. Inovasi Mandiri Pratama for providing early professional experience, which further strengthened my interest in pursuing advanced studies in Geomatics.

Finally, I would like to thank my parents and sisters for always encouraging me from afar. A very special thanks to my wife, Nuraulia Mugniza, for her constant love, support, and patience throughout this journey, especially during her pregnancy and the birth of our first child, which has meant the world to me. The arrival of our son, Aderra Ghaniyy Auliyaa Bainus, and his smile have been a source of strength and joy, brightening even the most challenging days.

Der Derian Auliyaa Bainus
Delft, June 2025

Contents

List of Figures

List of Tables

Acronyms

1 Introduction

1.1 Research Background and Motivation

High-resolution 3D digital representations of real-world objects are essential in geospatial analysis, enabling precise measurement and spatial understanding. Light Detection and Ranging (LiDAR) point cloud data plays a significant role in this field due to its ability to provide such detail. Advancements in various LiDAR post-processing techniques have enabled the effective use of heterogeneous point cloud data to capture dynamic environmental changes (Riofrío et al. 2022; Xiao 2012), making a seamless and unified dataset (Leusink 2024), or compensating for specific sensor limitations (Shao et al. 2022; X. Cheng et al. 2018). Heterogeneous point cloud data refers to point clouds collected at different times and with different sensors, often varying in resolution, accuracy, and attribute characteristics. These works emphasise the importance of accuracy and consistency in position when managing multiple heterogeneous LiDAR point clouds, ensuring that spatial discrepancies are minimised and that the datasets remain precisely aligned for reliable analysis. One technique used to address these spatial discrepancies is known as harmonisation.

Harmonisation is applied to correct spatial discrepancies in overlapping datasets, ensuring better accuracy and consistency in the positioning of heterogeneous point clouds. This process involves aligning datasets to minimise the datasets' misalignment caused by various sources of error. A key element in this process is the use of benchmarks, which serve as reference features to calculate positional differences between datasets. For a benchmark to function effectively, it must be detectable in both the reference dataset and the dataset being aligned, allowing for assessment and correction of spatial discrepancies.

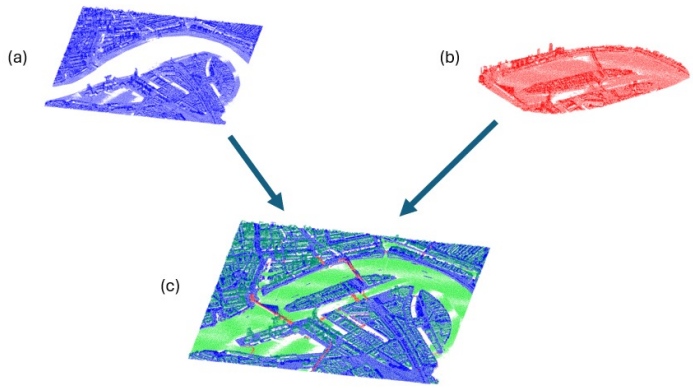


Figure 1.1: Illustration of harmonisation: (a) Reference point cloud. (b) Target point cloud. (c) Harmonisation result

In the Netherlands, an initiative called the Integrale Hoogtevoorziening Nederland (IHN) was launched at the Actueel Hoogtebestand Nederland (AHN) and Beeldmateriaal Congress 2024, held on 16 October 2024 in De Mariënhof, Amersfoort (Actueel Hoogtebestand Nederland (AHN) Congress 2024). One of the main objectives of the initiative is to harmonise national point cloud datasets to ensure the positional accuracy of elevation data across the country. It also aims to explore and promote methods for assessing and improving the alignment of overlapping point cloud datasets acquired at different times, with different sensors, or by different organisations (Actueel Hoogtebestand Nederland (AHN) Congress 2024). As part of the objective, identifying a reliable and widely applicable benchmark for alignment has become important, particularly a benchmark that can be consistently detected across homogeneous point clouds.

One commonly used approach to ensure positional accuracy and consistency between overlapping datasets is the deployment of artificial targets as physical benchmarks (Urbančič et al. 2019; Liang et al. 2014; Yanmin Wang et al. 2014). The approach allows operators to determine the optimal placement of benchmarks and how they should be detected in LiDAR point cloud data based on their unique shape, reflectance, or colour to make sure of their detectability. However, there are several drawbacks to using artificial targets as benchmarks. In large-scale survey areas, deploying artificial targets can be highly expensive in terms of both labour and cost. Additionally, artificial targets must be placed before data acquisition to be recorded in the dataset, making this method suitable only for pre-planned surveys. As a result, the method is not suitable for datasets collected without prior coordination or planning. An alternative type of benchmark that can solve this limitation is needed.

To overcome the limitation of artificial targets as benchmarks, feature extraction techniques from LiDAR point cloud data have been studied over the past decades to establish reliable benchmarks for measuring positional inconsistencies and aligning overlapping point clouds (L. Cheng et al. 2018; Wu and H. Fan 2016; Shao et al. 2022; Xu et al. 2022). The process involves extracting segments of points into feature primitives such as lines, surfaces, or any other feature to calculate the roto-translational difference between the two tested datasets (L. Cheng et al. 2018). For example, (Wu and H. Fan 2016) and (X. Cheng et al. 2018) propose using building surfaces as control features for registering overlapping point cloud data, which model building edges and roofs to surfaces or lines, considering buildings as a robust benchmark due to their stable structure and salient edges. Additionally, buildings are widely distributed across urban areas, making them effective reference features. While effective in urban environments, this benchmark is highly dependent on the presence of buildings in the dataset. Consequently, it is unsuitable for areas without buildings, such as highways and bridges, where alternative benchmarks are required to ensure accurate registration. Therefore, there is a need to develop an alternative benchmark that remains effective in areas without buildings, such as highways and bridges, ensuring accurate alignment across diverse environments.

To address the limitations of existing benchmarks in areas where buildings are sparse or absent, the use of road markings has emerged as a potential alternative. As part of the IHN project, Daan van der Heide, representing the Ministry of Infrastructure and Waterways of the Netherlands (Rijkswaterstaat) in the IHN project and serving as my mentor in this research, is investigating the potential of road markings as an alternative benchmark for point cloud co-registration, due to their colour and material make them noticeable from point cloud data through intensity values (Heide 2024). The illustration of road marking detected from intensity value can be seen in Figure 1.2.

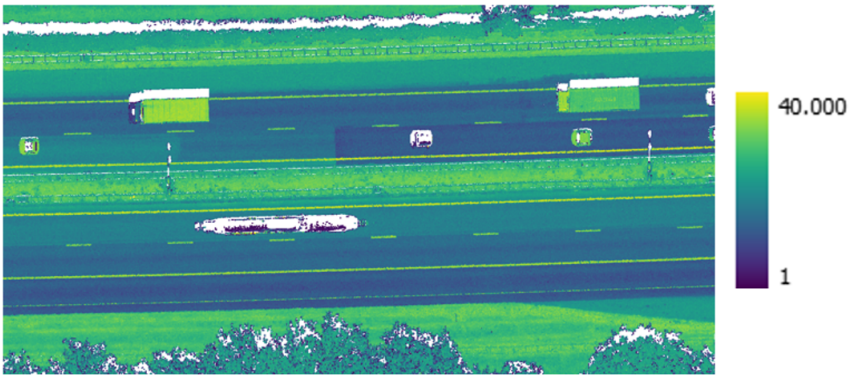


Figure 1.2: Road marking detected through point cloud intensity values

Road marking extraction techniques from LiDAR point cloud data have been studied over the past decade (Barçon et al. 2022), based on the assumption that painted road markings exhibit higher intensity values than the surrounding pavement. An approach to extract road markings from LiDAR point cloud data involves the use of machine learning and deep learning (Wen et al. 2019; Y.-T. Cheng et al. 2020). This method typically projects the 3D point cloud into a 2D image, enabling the model to learn the intensity patterns of road markings from labelled datasets. However, it requires a large amount of training data and has a high computational cost for implementation.

Another approach to road marking extraction is a rule-based method, which uses fixed mathematical rules rather than learning from data, as implemented in Yan et al. 2016. It works by identifying road points based on their height relative to the sensor, and detecting road markings by detecting sudden changes in intensity. However, this method is best suited for pre-planned surveys where the vehicle is specifically driven along the centreline to capture road markings accurately. This makes it less suitable for general LiDAR surveys, where the vehicle is not always positioned directly over the road markings.

Another method for extracting road markings is by filtering points based on an intensity value range, as applied in B. Yang et al. 2012. This method works by first removing points with intensity values outside a predefined range and retaining only those within it, then removing non-ground and noise points. The range is determined by analysing the typical intensity values of road markings in the specific point cloud data. However, in multi-source LiDAR datasets, intensity ranges can vary due to differences in sensors and preprocessing steps, making it difficult to define a fixed intensity range that can be reliably used across different datasets.

In (Heide 2024), a road marking extraction method is introduced that is intended to address the challenge of intensity variability across different LiDAR datasets. Unlike traditional fixed-threshold approaches, this method uses adaptive estimation of intensity thresholds suited to the intensity range of road markings within each segmented road surface. By dynamically filtering out non-road marking points, the method improves the reliability of road marking detection across heterogeneous datasets. Combined with geometric filtering and object clustering, the approach ensures that road markings can be modelled as 3D line-type benchmark representations.

While this method is designed to improve the reliability of road marking extraction across heterogeneous LiDAR datasets, the impact of its extraction accuracy on co-registration performance has not yet been evaluated. This highlights the need to assess not only how well the method performs, but also whether the current accuracy of extracted road markings is suitable for use in co-registration.

To address the gap, this research investigates how the extraction accuracy of the road marking extraction method proposed in (Heide 2024) influences the alignment of overlapping point clouds in the harmonisation of heterogeneous LiDAR data. The findings will provide insights into the feasibility of using the method to extract road markings as benchmarks, and into their impact on the alignment accuracy of LiDAR datasets.

1.2 Research Objective

This research aims to demonstrate the use of road markings extracted using the method proposed in Heide 2024 as a benchmark in the LiDAR point cloud harmonisation process and investigate how the extraction accuracy affects the alignment accuracy of LiDAR point cloud. To achieve this, the following main research question is proposed:

"How suitable are road markings extracted using the method proposed in Heide 2024 as a benchmark for heterogeneous LiDAR point cloud harmonisation?"

In this research, suitability refers to how well the extracted road markings fulfil their purpose as a co-registration benchmark. This includes examining whether they remain usable to reduce the misalignment distance between the overlapping datasets, given the current level of extraction accuracy, and understanding how the inaccuracies in road marking extraction may affect the alignment result.

To support the main research question, several sub-research questions have been formulated to ensure a focused approach toward the main research question, as listed below:

- How can reliable correspondences between road markings be established across LiDAR datasets with different acquisition times?
- How does weighting the road markings based on their geometric stability influence the accuracy of multi-temporal LiDAR co-registration?
- How does the accuracy of automatically extracted road markings influence the accuracy of LiDAR co-registration?
- How does the alignment error change with increasing distance from the area where road markings are used as a benchmark?
- What are the types of uncertainty that affect the accuracy of road marking extraction for position harmonisation?

1.3 Research Scope

The research focuses specifically on road markings extracted from LiDAR point clouds using intensity values as benchmarks. While other methods for generating point cloud data exist, such as photogrammetry-based point cloud reconstruction, which may still detect road markings using RGB values, these datasets generally lack intensity information and are thus excluded from this research.

The research also concentrates on the geometric aspect of point cloud harmonisation, specifically on positional inconsistencies between heterogeneous LiDAR datasets. It does not include harmonisation of intensity or RGB values, as the main goal is to achieve accurate spatial alignment rather than consistency in appearance or colour.

The study focuses on road marking data as the final product of the extraction method proposed in Heide 2024. The road marking extraction was carried out and provided by Rijkswaterstaat. However, a detailed examination of the extracted data is required for specific parts of the analysis. The data was made available exclusively for this research.

While various types of road markings on different road types can be detected in Dutch point cloud datasets, this study focuses specifically on simple dashed lane markings and block dashed lane markings found on highways, as shown in Figure 1.3. This choice reflects the current stage of development in road marking extraction methods, which are still developing.

Figure 1.3: Road mark types: (a) Dashed lane marking. (b) Block dashed lane marking

1.4 Thesis Outline

This thesis is structured into five main chapters, each addressing different aspects of the research. The outline of the thesis is as follows:

- **Chapter 1: Introduction**
  This chapter provides an overview of the research, including the background and motivation, research objective and research questions. It presents the reasons underlying the need to explore the implementation of the automatically extracted road markings as a co-registration benchmark.

- **Chapter 2: Related Work**
  This chapter reviews studies that are related to this research. It starts by reviewing sources that cause point cloud misalignment. Then, it reviews existing methods of road marking extraction from LiDAR data and how to validate them. After that, it reviews methods to align LiDAR point cloud data and methods to validate alignment accuracy.
- **Chapter 3: Methodology**
  This chapter details the research methodology, including the data sources, the proposed approach used to quantify road marks extraction quality, and the proposed methods to co-register point cloud data using the extracted benchmark.
- **Chapter 4: Results**
  This chapter presents the results obtained through the proposed approach. It analyses the outcomes and findings related to the implementation of automatically extracted road markings as a co-registration benchmark.
- **Chapter 5: Discussion**
  This chapter discusses the limitations found during the study. It discusses the challenges related to the study area, challenges in the processing step, and limitations of the proposed evaluation approach in this research.
- **Chapter 6: Conclusion and Future Work**
  The final chapter summarises the key findings of the study and provides recommendations for future research in implementing automatically extracted road markings as a co-registration benchmark.

1.5 Footnotes

The use of AI as writing assistance is acknowledged in this thesis to improve clarity and correct grammatical errors, under the full supervision of the writer. All research content, analysis, and conclusions are the sole responsibility of the writer.

2 Related Works

This chapter reviews several studies as the starting point of this research. It first explores LiDAR point clouds acquisition technique and the causes of errors. It then explores methods for extracting road markings from LiDAR data and ways to evaluate the extraction result. Followed by a review of co-registration methods for aligning LiDAR point clouds and approaches to evaluate point cloud alignment.

2.1 LiDAR Data Acquisition Techniques and Error Sources

Over the past decades, various LiDAR point cloud acquisition techniques have been developed to meet the growing demand for high-precision spatial data. Three of the most well-known techniques are Airborne Laser Scanning (ALS), Mobile Laser Scanning (MLS), and Terrestiral Laser Scanning (TLS). These techniques have emerged in response to various needs, such as achieving optimal coverage, resolution, and accuracy in different environments, as well as other factors like cost, platform availability, and technological constraints. Each system is designed to address specific challenges in LiDAR data acquisition by operating under different conditions, viewing target objects from varying perspectives, and relying on different positioning mechanisms. This section examines ALS, MLS, and TLS systems, along with the potential sources of error that can contribute to point cloud misalignment.

2.1.1 Airborne Laser Scanning

ALS systems can efficiently capture large-scale point cloud data using aeroplanes, helicopters or Unmanned Air Vehicle (UAV), making this technique ideal for environmental mapping. However, due to the high-altitude, top-down perspective, ALS datasets often have lower point density than those captured using MLS and TLS, limiting data collection in enclosed or obstructed areas Shao et al. 2022.

ALS conceptually records point cloud data by transferring the carrier's coordinates, obtained via Global Positioning System (GPS) integrated in the platform, to points on the Earth's surface. This process is achieved by measuring the laser pulse travel time and transmission angle, which are corrected using the carrier's flight attitude data (Killinger 2014). The technology integrates three key subsystems: Global Navigation Satellite System (GNSS), Inertial Navigation System (INS), and laser ranging technologies, each operating in its own reference frame as shown in Figure 2.1. However, the integration of these independent subsystems introduces systematic errors, primarily due to sensor calibration inaccuracies, time synchronisation issues, and intrinsic errors within each subsystem, leading to inconsistencies in the recorded point cloud Schenk 2001.

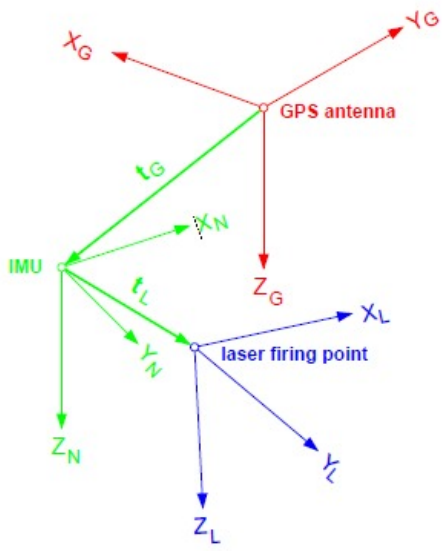


Figure 2.1: Airborne laser scanner subsystem reference frame Schenk 2001

Systematic errors in ALS can be categorised into several types, including range errors (biases in laser pulse travel time measurements), scan angle errors (misalignment in laser angles), INS drift errors (gyroscope and accelerometer inaccuracies), GNSS positioning errors (multipath effects and atmospheric delays), and mounting misalignments (incorrect sensor orientation relative to the aircraft) Schenk 2001; Xiaohong and Jingnan 2004. According to Schenk 2001, these errors propagate through the data acquisition process, introducing positional inconsistencies in overlapping flight strips, often leading to misalignment in ALS point clouds. Similarly, Xiaohong and Jingnan 2004 points out that GNSS, INS, and LiDAR integration errors can cause noticeable positional shifts in the data. These errors can become more significant depending on factors such as flight altitude, scan angle, and the nature of the terrain.

External factors such as changes in weather conditions and terrain variability can further increase these errors. Fog, rain, and snow can weaken the laser pulse, making distance measurements less accurate. Changes in ground height and thick vegetation can also affect how the laser reaches the surface and bounces back, leading to errors in elevation data Donkers 2024.

2.1.2 Mobile Laser Scanning

MLS system is a high-resolution 3D data acquisition technique that captures point cloud data while mounted on a ground vehicle such as a car or train. Unlike ALS, which operates from high altitudes, MLS offers a balance between spatial coverage and point density, making it particularly well-suited for applications in urban environments, transportation networks, and large-scale infrastructure projects Yanjun Wang et al. 2019. In contrast to ALS, which captures data from a top-down perspective, MLS operates closer to the ground, allowing for the detailed scanning of objects that may be obstructed in aerial views, such as building

facades and roadsides. This proximity results in denser point clouds and improved detail in feature representation.

Similar to ALS, an MLS system typically consists of a LiDAR scanner and a Position and Orientation System (POS), which is an integrated system comprising an Inertial Measurement Unit (IMU) and a GPS. These components work together to georeference the collected point cloud data Yanjun Wang et al. 2019; Shahraji, Larouche, and Cocard 2020, see Figure 2.2 for the location of subsystems in MLS system. However, because MLS operates near the ground in busy environments such as cities and constructed areas, GPS signals are more susceptible to multipath errors Kalenjuk and Lienhart 2022. In urban areas, GPS signals are often reflected off buildings, trees, and other obstacles before reaching the receiver, leading to inaccuracies in positioning. This increased susceptibility to multipath effects makes GPS errors more influential in MLS than ALS, where the sensor operates at higher altitudes with fewer obstructions.

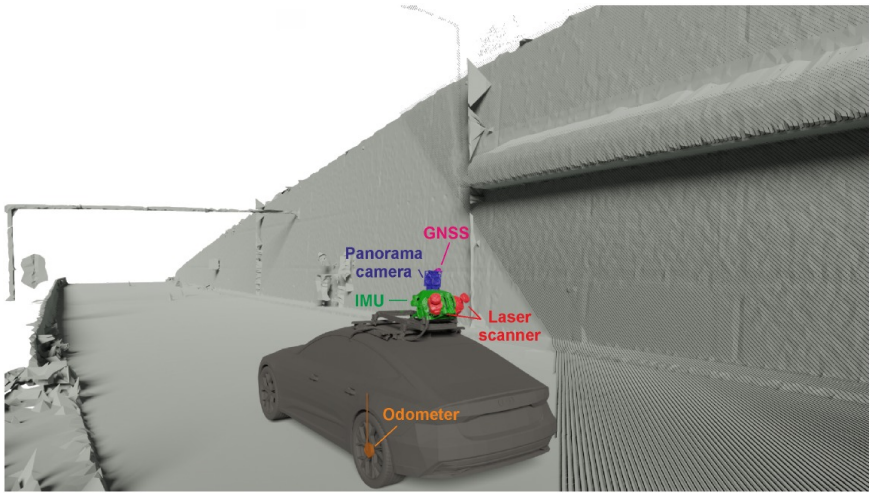


Figure 2.2: Mobile laser scanner system components (Kalenjuk and Lienhart 2022)

Additionally, MLS is prone to other systematic errors affecting georeferenced point clouds' accuracy (Shahraji, Larouche, and Cocard 2020; Kalenjuk and Lienhart 2022). According to Shahraji, Larouche, and Cocard 2020, a major source of error in MLS is misalignment between LiDAR sensor, IMU sensor, and GPS receiver, where discrepancies in roll, pitch, and yaw angles cause distortions, especially on inclined or vertical surfaces. Further mentioned in the study, leverarm offsets, which are the spatial distance between the LiDAR scanner and GPS or IMU can lead to systematic shifts, affecting georeferencing accuracy. Another source of errors mentioned is errors in distance measurement (range errors) and scanning angles, which can cause points to appear at the wrong position, especially on shiny or reflective surfaces, because the laser pulse may scatter, be partially absorbed, or reflect multiple times before returning to the sensor. Kalenjuk and Lienhart 2022 also mentions that external factors, such as scanning geometry, surface roughness, and reflectivity, influence the quality of the resulting point clouds.

2.1.3 Terrestrial Laser Scanner

TLS is a data acquisition technique capable of capturing highly detailed environmental conditions, producing the densest point cloud compared to other LiDAR techniques mentioned above Rodríguez-Gonzálvez et al. 2017; Kushwaha et al. 2023. This is due to its acquisition method, which utilises a stable platform, such as a tripod, positioned in a fixed location. TLS is particularly effective for modelling high-polygon objects with complex structures, such as trees, bridge frameworks, and cultural heritage sites, with exceptional detail. However, as TLS requires the scanner to be placed at specific points for data collection, acquiring large-area or large-object datasets takes significantly longer than other methods.

TLS sensor emits a laser pulse, splitting it into a reference and a scanning beam directed by a rotating mirror. The pulse reflects off the object, and the scanner calculates distance based on the time difference between the reference and received signals, generating a dense 3D point cloud Soudarissanane 2016 (see Figure 2.3).

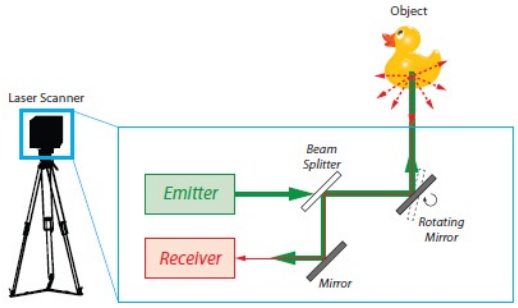


Figure 2.3: Terrestrial laser scanning mechanism Soudarissanane 2016

Unlike ALS and MLS, TLS does not rely on GNSS for positioning, as it operates from a static location rather than a moving platform. Instead, TLS typically uses ground control points or total station measurements for georeferencing point clouds. Another method of georeferencing TLS data is to position the scanner on a pre-measured fixed reference point during measurement. This fixed reference point is usually determined using the static GNSS method, which provides higher accuracy than kinematic methods, such as those used in ALS and MLS. Static GNSS positioning improves accuracy due to longer observation durations, allowing for better error filtering and averaging, whereas kinematic GNSS computes positions in real-time at each measurement epoch, making it more susceptible to instantaneous errors and dynamic model constraints Tiberius et al. 2021.

Soudarissanane 2016 mentioned several factors influencing individual TLS points' quality, which are: systematic and random errors from the scanner-related errors, environmental conditions, and object surface properties. Scanner-related errors, such as beam divergence, uncertainty in the angular position of the rotating mirror, and axis misalignment between laser beam, rotating mirror, and the scanner's internal axis, introduce random errors. Environmental conditions, including dust, temperature, and humidity, can degrade the laser signal. The surface properties of scanned objects, such as reflectivity and roughness, affect the intensity and number of valid returns. Finally, scanning geometry, particularly the incidence angle and range, directly impacts measurement precision.

2.1.4 Overview of LiDAR Acquisition Techniques

To summarise the subsection, the characteristics and potential error sources of various LiDAR acquisition methods are compared. This includes airborne, mobile, and terrestrial laser scanning systems, each with distinct positioning strategies, sensor configurations, and operational environments. The following tables provide an overview of their key components, positioning methods, and typical sources of error. This structured comparison highlights how different acquisition techniques may introduce specific limitations in point cloud quality, understanding the impact on applications such as co-registration.

Table 2.1: LiDAR acquisition types with positioning methods and components

Acquisition Type	**Positioning Method**	**Component**
Airborne Laser Scanning	Kinematic (airborne)	GNSS, INS, LiDAR scanner
Mobile Laser Scanning	Kinematic (on ground)	GNSS, INS, LiDAR scanner, odometer
Terrestrial Laser Scanning	Static (on ground)	GNSS, LiDAR scanner

Table 2.1 outlines the positioning methods and system components of each LiDAR acquisition technique. On the other hand, Table 2.2 builds upon this overview by presenting a detailed comparison of the primary error sources associated with each system. These errors are grouped into categories, which are: positioning, mounting, scanning, environmental, and surface interaction, showing how different types of LiDAR systems are affected by specific technical and environmental factors.

Table 2.2: Comparison of LiDAR acquisition methods and error sources

Acquisition Type	**Positioning Error**	**Mounting / Integration Error**	**Scanner Error**	**Environment Factor Error**	**Surface Interaction Error**
Airborne Laser Scanning	GNSS multipath, INS drift	Lever-arm offset, POS–LiDAR misalignment	Range error, scan angle error	Sensitive to weather (fog, rain, snow)	Reflective surfaces (e.g. water, metal)
Mobile Laser Scanning	GNSS multipath, INS drift, odometer error	Lever-arm offset, POS–LiDAR misalignment	Range error, scan angle error	Affected by urban obstructions	Reflective surfaces (e.g. water, metal)
Terrestrial Laser Scanning	Minimal (static GNSS or total station)	Rotating mirror angle errors, axis misalignment	Beam divergence, range error, scan angle error	Dust, temperature, and humidity	Reflective & roughness of texture surfaces

2.2 Road Marking Extraction from LiDAR Point Cloud

Road markings are among the most detectable and geometrically distinct features in urban environments, making them useful not only for applications like high-precision map creation and autonomous driving (Wen et al. 2019; R. Yang et al. 2020; Yao et al. 2021), but also as potential geometric benchmarks for point cloud alignment (Heide 2024). The extraction of road marks from LiDAR data, however, is a challenging task and depends on how the road marks detected in point cloud data through intensity value, influenced by surface reflectivity, the condition of road marks at the data acquisition phase and the characteristics of the LiDAR sensor used.

Many existing road marking extraction methods rely on the assumption that painted road markings exhibit higher reflectivity than the surrounding pavement. Intensity values recorded by LiDAR sensors capture this contrast, enabling various approaches to utilise intensity information for detecting road markings within point cloud data.

Machine learning and deep learning methods have gained popularity for road marking extraction due to their ability to learn complex spatial and reflectance patterns from LiDAR data. These approaches typically rely on supervised learning, where labelled datasets are used to train models, with LiDAR intensity serving as a key feature. A common strategy involves projecting the 3D point cloud into a 2D grid and using the average intensity in each cell as input for segmentation. This enables deep learning models to learn reflectance patterns associated with road markings and achieve robust performance across different pavement types and intensity distributions (Wen et al. 2019; Y.-T. Cheng et al. 2020). Although deep learning models can generalise well when trained on diverse datasets, they require large amounts of labelled data and high computational resources.

Another approach to extracting road markings involves traditional model-based and rule-based methods, which use fixed mathematical rules rather than learning from data. Model-based methods fit simple shapes, such as lines or curves, to the data, while rule-based methods rely on features like intensity differences or spatial patterns. For example, Yan et al. 2016 proposed a method that processes 3D mobile LiDAR data by identifying the road surface based on height differences and detecting markings through sudden changes in intensity along scan lines. Similarly, Mi et al. 2021 used a two-step method: first selecting possible road marking points using intensity and elevation information, then improving the results by fitting simple geometric models to better match the shape of the markings.

Intensity-based filtering approaches make use of the characteristic intensity range of road markings within point cloud datasets. A straightforward method for filtering road marking points using intensity value ranges was proposed by B. Yang et al. 2012. The method works by applying a two-step filtering process: the first step filters road points with intensity values within the fixed range of 100 to 300, while the second step further refines the filtered points based on their elevation. This is done to eliminate objects with similar intensity values to road markings, such as vehicles and other road features.

While effective under controlled conditions, these approaches are highly sensitive to various factors, including sensor type, scanning geometry, and environmental conditions. For instance, the angle of incidence between the laser beam and the road surface can influence the return intensity, as can the range to the object and the surface moisture (Li et al. 2023; Ackroyd et al. 2024). Furthermore, the intensity scale is not standardised across different LiDAR systems or LiDAR acquisition projects (see Figure 2.5), and even within the same system, it can vary depending on calibration settings or operating conditions. Consequently,

using fixed range values may fail to filter road marking across datasets captured by different LiDAR systems or projects.

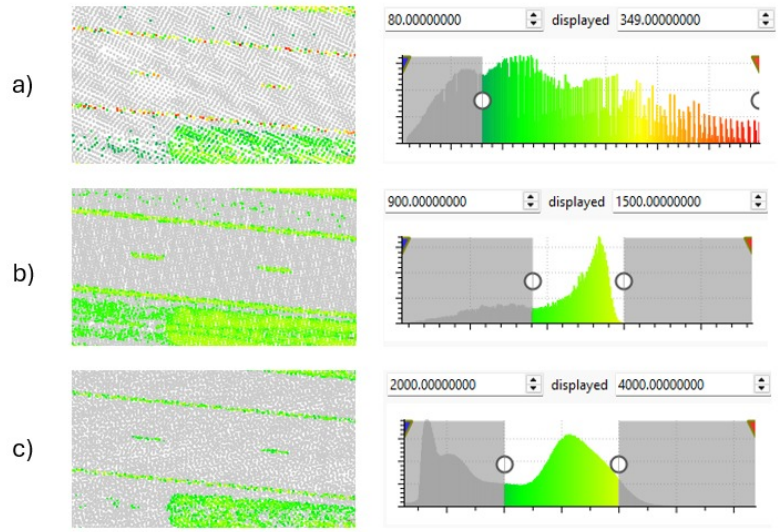


Figure 2.4: Variation of road marking intensity ranges across datasets from multiple projects: a) AHN3 dataset (2018), road marking detected in range 80 to 349. b) AHN4 dataset (2022), road marking detected in range 900 to 1500. c) AHN5 dataset (2024), road marking detected in range 2000 to 4000

To solve the limitation of road marking extraction using the fixed intensity range filter approach as mentioned above, an adaptive intensity range filter, which is capable of automatically determining the threshold based on the input point cloud data, should be employed.

Yao et al. 2021 proposed a method for segmenting road marking point clouds based on intensity by first rasterising 3D road point clouds into 2D intensity images. An adaptive threshold technique is then applied to the raster image, whereby the binarisation threshold for each pixel is determined by the local intensity distribution of its surrounding neighbourhood. Points within each pixel with an intensity value above the mean of their neighbourhood are then extracted as road marking points. The approach enables dynamic threshold adjustment across datasets, accommodating variations in data acquisition techniques or post-processing steps. However, the method requires the transformation of 3D point clouds into 2D images for road marking extraction, which results in the loss of elevation information. For co-registration purposes, where 3D information is essential, an alternative approach is needed that can extract road markings as 3D features.

A method that uses an adaptive intensity filtering approach to extract road marking as a 3D feature was introduced by Heide 2024. It was designed to extract road markings from LiDAR data to be used as benchmarks for co-registration. Co-registration requires at least two overlapping point clouds, and these point clouds may differ in their intensity ranges. This variation makes the adaptive intensity threshold useful for consistently extracting road markings across different datasets. The pipeline of this method is shown in Table 2.3

Table 2.3: Processing pipeline of the automatic road marking extraction method using adaptive intensity threshold introduced in Heide 2024

Number	Step name
Step 0	Gridding and cleaning the input point cloud
Step 1	Road segmentation
Step 2	Curb and noise segmentation
Step 3	Adaptive estimation of the intensity threshold
Step 4	RM cluster estimation
Step 5	Line segmentation
Step 6	Zigzag- and direction correction
Step 7	Naming the line segments

The method works by first dividing the point cloud datasets into smaller tiles to make processing more manageable. Each tile is then cleaned and segmented using external road network data to isolate the road surface. ====ADD HERE=====. To identify road markings, an adaptive threshold is calculated per tile by analysing the shape of the intensity histogram, allowing the method to adjust to different datasets and acquisition conditions. The resulting threshold is smoothed to reduce noise and applied to extract high-intensity points, which are then grouped into clusters and filtered using simple geometric features, as listed in Table 2.4, to remove those that do not represent road markings.

Table 2.4: Geometric features and formulas as parameters to filter road marks cluster proposed in Heide 2024

Name	Formula
Linearity	$\frac{\lambda_1-\lambda_2}{\lambda_1}$
Planarity	$\frac{\lambda_2-\lambda_3}{\lambda_1}$
Sphericity	$\frac{\lambda_3}{\lambda_1}$
Anisotropy	$\frac{\lambda_1-\lambda_3}{\lambda_1}$
Sum of the Eigenvalues	$\sum_{i=1}^{3}\lambda_i = \lambda_1+\lambda_2+\lambda_3$
Change in Curvature	$\frac{\lambda_3}{\sum_{i=1}^{3}\lambda_i}$

The remaining clusters are then converted into line features using Principal Component Analysis (PCA) and convex hull fitting. The method addresses inconsistencies in marking direction by identifying and correcting zigzag patterns that arise from misaligned clusters. It first detects breakpoints in the line segments and compares their orientation to the local road direction using data from the Dutch National Road Database (NWB). If a segment deviates by more than a determined value, it is labelled as zigzag and removed. Lastly, unique identifiers are assigned to the corrected line segments.

2.3 Evaluation of Automatic Object Extraction Method

The quality of the automatic road marking extraction method used in this research is an essential factor in evaluating whether extraction accuracy affects point cloud alignment results. Therefore, this section reviews the methods used to evaluate the quality of automatic object extraction as a starting point for assessing the accuracy of the automatic road marking extraction method used in this research.

Heipke et al. 1997 proposed a framework for evaluating automatic road extraction algorithms by comparing the results against manually plotted road lines. Their method involves two steps: matching the automatically extracted road network to the reference data, and calculating its quality measure. The matching is done by placing a buffer around the reference road lines. Any extracted road segments that fall inside this buffer are counted as correctly matched, or true positives. The process is then repeated in the opposite direction by placing a buffer around the extracted roads to see which parts of the reference roads they overlap. This helps to detect false positives, or incorrect extractions, and false negatives or missed roads. This two-way matching provides a basis for measuring the accuracy and completeness of the extraction.

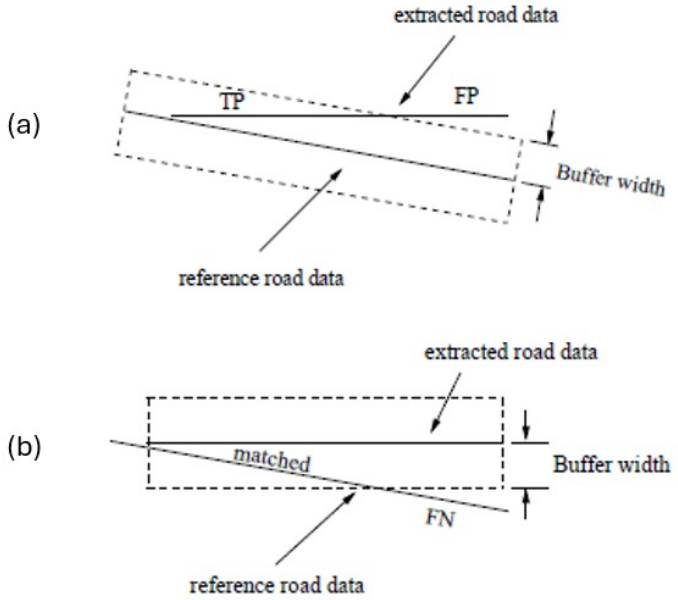


Figure 2.5: Matching principle defined in Heipke et al. 1997: a) Matched extraction. b) Matched reference.

After the matching step, the second part of the method proposed in Heipke et al. 1997 is to calculate quality measures that evaluate how well the road extraction performed. These measures include completeness, which tells how much of the reference road data was correctly found, and correctness, which shows how much of the extracted data is actually correct. A combined quality score considers both completeness and correctness. In addition, the Root Mean Square Error (RMSE) is used to quantify the average positional error between the extracted roads and the reference lines. This helps assess how accurate the extracted road

position is. Together, these metrics provide a quantitative evaluation of the extraction results and make it comparable to the other automatic road extraction methods.

Gao et al. 2017 presented a method to automatically extract pavement markings from mobile LiDAR data. To evaluate the accuracy of the extraction, they compared each detected marking to manually labelled reference data using two metrics: correctness and completeness. Correctness was calculated individually for each extracted marking by comparing its centre position to the centre of the corresponding reference marking. The closer the two centres were, the higher the correctness score for that marking. The final correctness score reported for a dataset was obtained by averaging the individual scores across all detected markings. Completeness was also calculated for each extracted marking by comparing the area covered by the extracted marking to the area of the corresponding reference marking. This measured how much of the actual marking was successfully captured. Like correctness, the final completeness score was the average of all individual completeness scores across the dataset.

2.4 Point Cloud Co-Registration

Position harmonisation aims to ensure that multi-source point cloud datasets are geometrically consistent, meaning that corresponding features occupy the same relative positions and align accurately in 3D space within a shared spatial reference (Riofrío et al. 2022). A key process within position harmonisation is co-registration, which is a process to calculate the roto-translational differences between the dataset to be aligned and the reference, refining the alignment of two or more point clouds that are already in the same coordinate system. Therefore, this section will provide a brief overview of the co-registration process.

Methods to align two overlapping datasets have been widely studied over the past decades (X. Huang et al. 2021; L. Cheng et al. 2018; Fontana et al. 2021; Brightman, L. Fan, and Zhao 2023). The progress has been driven by the challenge of misalignment in overlapping point cloud data collected using various acquisition techniques across different survey locations. Research in this area has focused on developing reliable methods for each step of the process, including extracting key features from point cloud data, establishing correspondences between overlapping datasets, and computing roto-translational differences based on these correspondences (Wu and H. Fan 2016; Shao et al. 2022; Xu et al. 2022).

2.4.1 Feature Extraction

Feature extraction algorithms in point cloud registration are often developed considering the specific characteristics of the study area, as certain environments contain only particular types of objects that can be detected or modelled. The following are examples of feature extraction used for point cloud registration.

In Xu et al. 2022, where the overlapping features are underwater point clouds acquired by airborne LiDAR bathymetry, it is challenging to identify solid objects on the seafloor that can serve as key features for registration. Therefore, the study uses the curved lines representing seafloor topographic features to evaluate the alignment between the two overlapping point clouds. To extract these curve features, the underwater points are first used to generate isolines. Then, a cubic parabolic spline curve interpolation is applied to enhance the

smoothness and continuity of the isolines to enable better correspondence with the matching curves in the overlapping dataset. Based on this approach, it is assumed that the geometric quality of the extracted key features may play a significant role in the registration process. However, this aspect is not explored in detail within the study.

In Shao et al. 2022, the study area is located in a forest environment, where the scene is mainly composed of trees and contains few ground objects that can be used as key features for the coarse alignment process. Therefore, the study proposes a method to extract key points from the shapes of tree canopies. To achieve this, the point cloud data are first projected onto 2D top-view images to better represent the canopy shapes, then corners with significant geometric gradient changes on canopy contours are extracted as key features by determining the connectivity between each point on the curve and its four neighbouring points. Prior to key point extraction, image filtering is applied to address noise, holes, and irregular canopy edges caused by the discreteness of LiDAR points. The image filtering process conducted is believed to ensure that the extracted key points are reliable and suitable for use in the alignment process. However, the study does not provide a detailed discussion on the extent to which reliable key features are required for effective point cloud alignment.

2.4.2 Corresponence Establishment

Correspondence establishment means finding matching features between two overlapping point clouds. The step is crucial in the co-registration process because it allows the alignment of the datasets by estimating how one should be rotated and shifted to match the other. The type of features used for matching can vary depending on the data and the environment. Some methods use geometric features like curves or key points, while others use neighbourhood-based comparisons such as matching by grids. The following examples show different ways to establish correspondences based on the characteristics of the data and the size of the misalignment.

To establish correspondences within the overlapping region in Xu et al. 2022, the study employs the Longest Common Subsequence (LCSS) algorithm in combination with a curve deformation energy function to identify matching curves between the two point clouds. This approach demonstrates how extracted curve line features can be effectively utilised to compute positional differences and support the alignment of overlapping datasets. The key challenge in this process lies in automatically identifying reliable correspondences when a significant spatial gap exists between the datasets, which can complicate curve matching, as illustrated in Figure 2.6. In contrast, when the positional discrepancies are minor and the misalignment is relatively small, establishing correspondences becomes more straightforward and can often be achieved using simpler matching techniques.

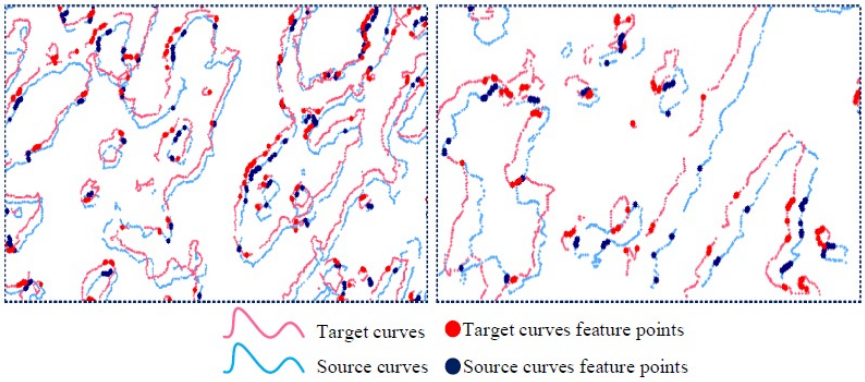


Figure 2.6: Illustration of gap between feature points (Xu et al. 2022)

Unlike the approach in Xu et al. 2022, which addresses significant gaps between overlapping datasets, this method makes use of the relatively small gap between the two datasets to perform grid analysis to establish correspondences. Due to the small gap between feature pairs in each correspondence, the method uses a straightforward spatial distance-based approach to establish correspondences, rather than relying on the computation of feature descriptors to identify matches. By employing a fixed grid size, it aligns overlapped point cloud data by assessing the similarity between corresponding points from both datasets and uses a fixed threshold of similarity to filter out low correspondences as shown in Figure 2.7. This approach ensures that any horizontal misalignment stays within the predefined grid boundaries, thereby facilitating the accurate establishment of correspondences.

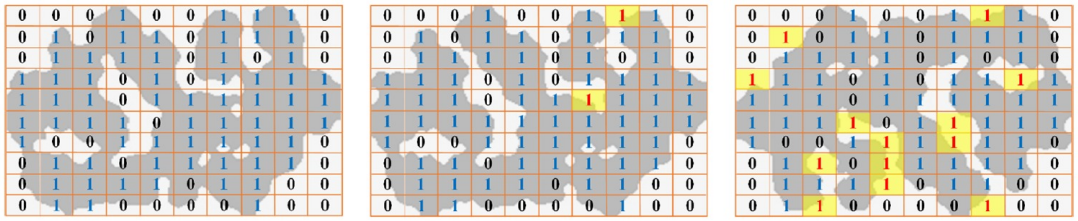


Figure 2.7: Illustration correspondence establishment approach proposed in Shao et al. 2022. Yellow grid indicates incorrect correspondence; (a) matched image by gridding, (b) Filtering incorrect correspondence in each cell using a fixed threshold, (c) filtering incorrect correspondence in each cell using higher threshold value

2.4.3 Computation of Rotation and Translation

The rotation and translation parameters computation between datasets is a fundamental step in co-registration. The process involves using corresponding features from both datasets to estimate how one should be rotated and shifted to align with the other. The type of features used, such as points, lines, or planes, often determines the method used to compute the transformation, as different feature types require different mathematical approaches. The following examples explain how rotation and translation are computed.

One well-known method for estimating rigid transformations using point-based features is presented in B. K. P. Horn 1987. The method operates on two sets of 3D points that represent the same object in different positions. Begins by calculating the centroid of each point set and shifting both sets so that their centroids align at the origin. The method then compares corresponding points and calculates the optimal rotation using quaternions, which are commonly used to represent 3D rotations due to their numerical stability. Once the rotation is determined, the translation is computed as the difference between the centroids after applying the rotation. Another alternative was proposed in Arun, T. S. Huang, and Blostein 1987; Umeyama 1991. Instead of quaternions, this method employs a least-squares formulation and solves it using Singular Value Decomposition (SVD) to find the optimal rotation. The translation is then derived in a similar way, by comparing the shifted centroids. The process is illustrated in Figure 2.8.

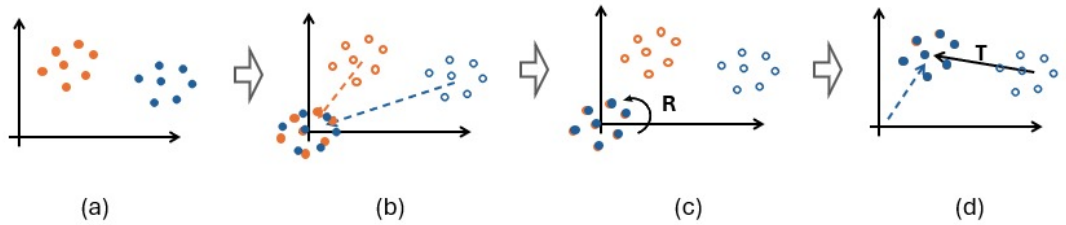


Figure 2.8: Illustration of transformation value calculation. Orange points are reference, blue points are target; (a) Untransformed condition. (b) Centroid alignment. (c) Rotation calculation. (d) Translation calculation.

In Arun, T. S. Huang, and Blostein 1987, the authors describe a method to find the best-fit transformation between two sets of 3D points using least squares. The objective is to compute a rigid transformation, comprising a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$, that best aligns the target dataset to the reference dataset using least-squares. Let $\mathbf{P}$ denote the set of points from the target dataset, and $\mathbf{Q}$ denote the corresponding set of points from the reference dataset. Each pair $(\mathbf{p}_i, \mathbf{q}_i)$ is assumed to be a known correspondence. The goal is to find the rigid transformation such that the transformed target points $\mathbf{R}\mathbf{p}_i + \mathbf{t}$ are as close as possible to the reference points $\mathbf{q}_i$.

This objective is formalised by minimising the sum of squared Euclidean distances between the transformed target points and the reference points:

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} \|\mathbf{q}_i - (\mathbf{R}\mathbf{p}_i + \mathbf{t})\|^2 \tag{2.1}$$

The process begins by computing the centroids of both point sets:

$$\bar{\mathbf{p}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{p}_i, \quad \bar{\mathbf{q}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{q}_i \tag{2.2}$$

Each set is then centralised by subtracting its centroid from all its points:

$$\mathbf{p}_i' = \mathbf{p}_i - \bar{\mathbf{p}}, \quad \mathbf{q}_i' = \mathbf{q}_i - \bar{\mathbf{q}} \tag{2.3}$$

The cross-covariance matrix $\mathbf{H}$ between the centralised point sets is computed as:

$$\mathbf{H} = \sum_{i=1}^{N} \mathbf{p}_i' \mathbf{q}_i'^T \tag{2.4}$$

To solve for the optimal rotation matrix $\mathbf{R}$, the Singular Value Decomposition (SVD) of $\mathbf{H}$ is performed:

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{2.5}$$

The optimal rotation matrix that minimises the least-squares error is given by:

$$\mathbf{R} = \mathbf{V}\mathbf{U}^T \tag{2.6}$$

However, if the determinant of $\mathbf{R}$ is negative (i.e., $\det(\mathbf{R}) < 0$), it indicates a reflection rather than a proper rotation. In this case, the last column of $\mathbf{V}$ is negated to ensure a proper rotation matrix:

$$\mathbf{V}[:,3] \leftarrow -\mathbf{V}[:,3], \quad \mathbf{R} = \mathbf{V}\mathbf{U}^T \tag{2.7}$$

Once the rotation matrix $\mathbf{R}$ is computed, the translation vector $\mathbf{t}$ is determined by aligning the centroids:

$$\mathbf{t} = \bar{\mathbf{q}} - \mathbf{R}\bar{\mathbf{p}} \tag{2.8}$$

This results in a rigid transformation $(\mathbf{R}, \mathbf{t})$ that can be applied to the target point set $\mathbf{P}$ to bring it into optimal alignment with the reference point set $\mathbf{Q}$.

In practice, point correspondences are often imperfect due to noise, occlusion, or partial overlap. To refine the alignment and improve accuracy, iterative methods such as Iterative Closest Point (ICP) are commonly used. This method was first introduced in Besl and McKay 1992 as part of research on estimating the optimal transformation matrix for point cloud alignment. ICP works by repeatedly updating the point correspondences based on the current alignment, then recalculating the transformation that best fits the updated pairs. This process continues until the changes become negligible, resulting in a more precise alignment between the point sets. Another widely used iterative method is Random Sample Consensus (RANSAC), which is designed to handle outliers and was first introduced in Fischler and

Bolles 1987. RANSAC randomly selects small subsets of point pairs to estimate a transformation, then evaluates how many other points agree with this transformation; these are referred to as inliers. This process is repeated many times, and the transformation with the highest number of inliers is selected. By focusing on the most consistent subsets, RANSAC increases robustness in situations where some correspondences are incorrect or affected by noise.

Another approach of rotation and translation matrix calculation, but based on plane features, is introduced in Wu and H. Fan 2016, who propose a method for aligning airborne LiDAR datasets using planar roof surfaces instead of individual points. In this method, flat roof areas are extracted from both datasets and represented by their normal vectors, which describe the orientation of each roof surface. The rotation is estimated by aligning the normal vectors of corresponding roof planes using a least squares solution. After rotation, the translation is computed by comparing the positions of the same planes in both datasets, specifically focusing on differences in their vertical placement. While this method uses different types of features, it shares the same overall goal of computing a rigid transformation to align two datasets.

Another feature-based approach is proposed in Habib et al. 2008, which performs strip adjustment using linear features extracted from the intersections of planar patches. These features represent consistent structures, such as roof ridges and surface edges, and are identified semi-automatically. Once corresponding lines are established, differences in their positions and orientations are used to calculate the transformation parameters required to align the strips. Specifically, changes in the direction and location of the lines are used to estimate the translation and rotation in 3D space. Line orientation information plays a critical role in determining the rotation parameters in this approach; therefore, the method is highly sensitive to orientation errors and inaccuracies in line angles.

2.5 LiDAR Point Cloud Accuracy Assessment

The term accuracy in LiDAR point cloud data is commonly categorised into two types: absolute and relative (Heide et al. 2024; Donkers 2024). Absolute accuracy refers to the positional discrepancies between the LiDAR data and its coordinate reference frame, which is represented by spatial measurements in the real world, commonly known as benchmarks. This metric reflects the correctness of the data with respect to reality or the actual situation. Relative accuracy, on the other hand, refers to the spatial agreement between overlapping LiDAR datasets and is defined by Rijkswaterstaat as a continuous value representing the distance between datasets acquired from two different measurement setups within the overlapping area (Heide et al. 2024). Illustration of absolute and relative accuracy shown in Figure 2.10. To effectively assess these accuracies, researchers develop various strategies that quantify the quality of the final product.

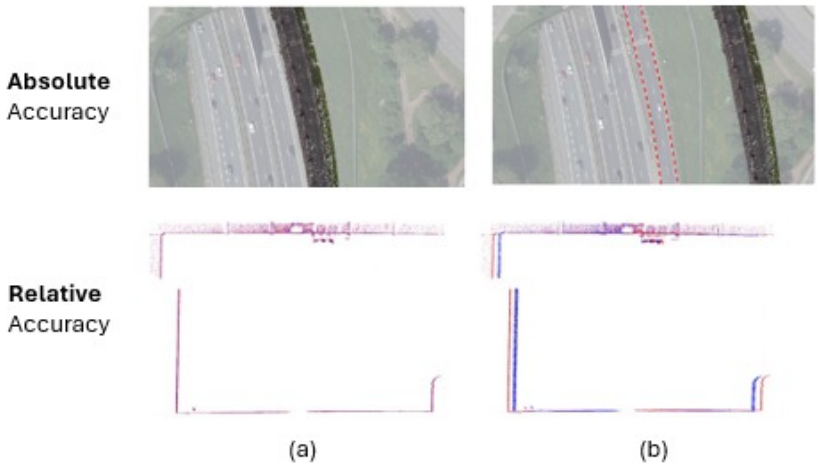


Figure 2.9: Absolute and relative accuracy described in Donkers 2024; (a) high accuracy, (b) low accuracy

2.5.1 Absolute Accuracy

In measuring LiDAR point cloud absolute accuracy. Csanyi and Toth 2007 utilises elevated disc-shaped targets as benchmarks. These benchmarks are specifically designed, both in colour and elevation, to be detectable in the point cloud data based on their intensity and height values. The centre of each benchmark is measured using GPS with centimetre-level accuracy. This enables the absolute accuracy of the LiDAR data to be assessed by comparing the coordinates of the benchmark centres derived from the GPS measurements with those extracted from the LiDAR point cloud. However, this approach requires additional techniques to accurately measure the benchmark coordinates, such as the use of GPS, which was implemented in this study. This makes the approach more suitable for controlled or planned surveys, where such equipment and preparation can be arranged in advance.

Another type of benchmark, tested in Donkers 2024. The research uses spherical targets distributed throughout the measurement area to assess the absolute accuracy of a LiDAR point cloud. These targets are detectable based on their shape and reflectivity in the intensity values. The assessment involves extracting the spherical targets from the point cloud, estimating the centres of the spheres, and comparing these coordinates with the corresponding reference values. The approach enables the centre of the target's coordinates to be estimated directly from the point cloud data, allowing for the assessment of absolute accuracy. However, the targets must be placed in the area prior to LiDAR data acquisition, making this method more suitable for planned surveys than for applications involving previously acquired datasets.

In addition to artificial targets, building features such as ridge lines are used as benchmarks to measure the absolute accuracy of point cloud datasets. A ridge line is defined as the horizontal 3D intersection between two sloped roof surfaces with opposing inclinations. Keuris 2024 introduced two approaches for building ridge line extraction: the direct approach and the indirect approach. The direct approach extracts ridge lines solely from the classified point cloud data, without prior knowledge of roof geometry. It operates by dividing the

point cloud into a grid, estimating the slope and orientation of each cell, grouping adjacent cells with similar characteristics into roof planes, and identifying ridge lines as intersections between pairs of roof planes with opposing slopes. On the other hand, the indirect approach extracts ridge lines using predefined roof surface geometries as input. It uses these known surfaces to directly select nearby points in the point cloud and then re-estimates the two roof planes. The ridge line is then calculated as the intersection between the refined planes. There are several ways to compute the distance between lines mentioned in Keuris 2024, which are: the distance between the centres of the two lines, the projected distance from the centres of one ridge line to the other, the average distance between the endpoints of the two lines, the difference in orientation between the two ridge lines.

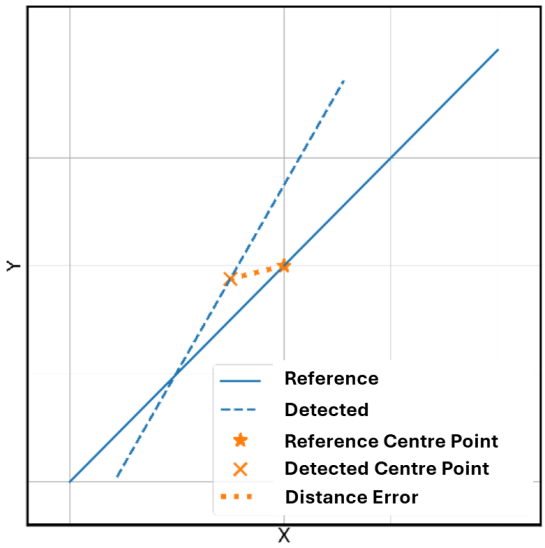


Figure 2.10: 2D representation of difference calculation between two ridge lines: using the distance between the centres of the two lines (Keuris 2024)

2.5.2 Relative Accuracy

Various approaches have been developed to assess the accuracy of point cloud data relatively. In Vosselman 2008, the assessment is performed by extracting ridge line point clouds from overlapping LiDAR strips and using them as surfaces for comparison. A straight line is fitted to each ridge in both strips, and the perpendicular distance between the corresponding lines is calculated. These distances represent the planimetric discrepancies between the strips and are used to evaluate the relative accuracy of the LiDAR data. However, the method relies on the presence of buildings, making it unsuitable for point cloud strips that do not contain building structures.

In Kim et al. 2022, the accuracy between point cloud datasets is measured using amorphous objects, or objects without clearly defined shape or form, such as trees and bushes, as the basis for comparison. Isolated objects are extracted from the point clouds, and an optimisation algorithm is applied to measure the positional differences between each segmented object and its corresponding match in the other dataset. However, since the objects used in the test are trees and bushes, which are sensitive to shape changes over time due to growth

or pruning, this method works best when both point cloud datasets are acquired within a short time interval.

Another approach is proposed in Donkers 2024, which assesses relative accuracy by extracting surfaces within overlapping regions, fitting a plane to each surface, and then comparing these planes with corresponding points at the same locations to measure the distances between the points and the fitted planes. In this approach, it is important to carefully select the region to ensure a valid assessment using this approach, as overly noisy areas can negatively affect the accuracy of the plane fitting. However, this method is tested in relatively small areas. In larger and more complex environments, such as urban areas where point cloud data can vary a lot due to diverse terrain and surface features, fitting a single plane to an entire region may result in high fitting errors and reduce the reliability of the accuracy assessment.

Another approach to measuring relative accuracy is through Digital Terrain Model (DTM) differencing, as demonstrated in Salach et al. 2018. This method involves filtering ground points from both the target and reference point clouds and converting them into rasterised DTMs. The target DTM is then subtracted from the target DTM to generate a difference raster, where each pixel represents the elevation error between the two datasets. This approach produces a large number of elevation differences that can be statistically analysed to provide a more robust estimate of relative accuracy. However, it assumes that horizontal errors are minimal and are effectively reduced by averaging elevations within each pixel, provided that the pixel size is larger than the expected horizontal error.

3 Methodology

This chapter will discuss the proposed methodology of the research to achieve the research objective defined in Chapter 1 and based on the related works reviewed in Chapter 2. The main workflow proposed in this research is illustrated in Figure 3.1.

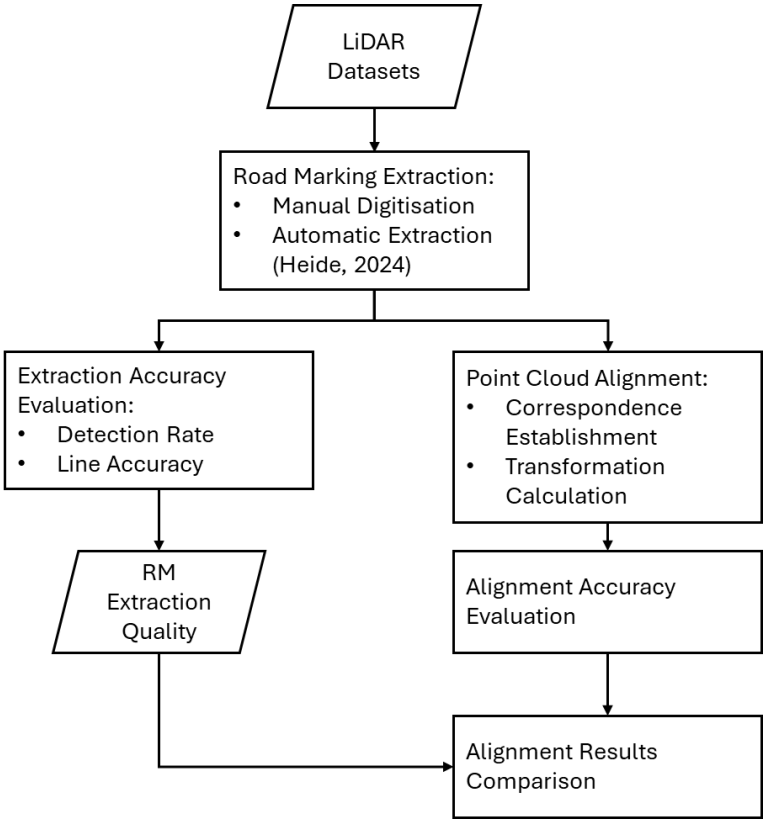


Figure 3.1: Workflow proposed in this research

The methodology of this research is structured around the use of road markings as benchmarks for the co-registration of heterogeneous LiDAR datasets. This study utilises point cloud datasets acquired from different sources and time periods, as described in Section 3.2. The process begins with the extraction of road markings from the selected datasets using the method proposed in Heide 2024. In parallel, manual digitisation of road markings is performed on the same datasets to create reference features, as detailed in Section 3.4. These manually digitised road markings serve as a reference to assess the geometric quality of the automatically extracted features, as explained in Section 3.5.

The RANSAC inlier voting approach is used to establish correspondences, where consistent matches between road markings across datasets are identified through repeated sampling. The process begins with an initial nearest neighbour search based on the 3D centroids of road markings within a defined spatial threshold, followed by an inlier-counting strategy to identify geometrically stable correspondences, which are described in Section 3.6. These correspondences are then used to estimate the rigid transformation parameters required for co-registration using the RANSAC-weighted centroid alignment, where each correspondence is weighted according to its inlier count described in Section 3.7.

In the final step of this research, the alignment results are compared to assess how the quality of road marking extraction affects co-registration accuracy. Alignments using automatically extracted road markings are compared with those using manually digitised ones, which serve as the reference. In addition, comparisons across different alignment time gaps are carried out to evaluate how well road markings perform as benchmarks over varying temporal differences.

3.1 Study Case

To provide a clearer focus for the research, a case study is conducted on aligning the ProRail–Spoorinbeeld point cloud with AHN data. The ProRail-Spoorinbeeld dataset is a digital twin of the Dutch railway network, capturing the condition of the railway and its surroundings, approximately 50 metres to the left and right, in multiple epochs. The dataset is part of the public point cloud inventory managed under the IHN project (Heide et al. 2024), making it a relevant subject of study. However, several sections of the railway data are located in areas isolated from buildings, where only highways and limited built structures are present 3.2. These conditions make such locations ideal for testing the potential of road markings along highways as alternative benchmarks, particularly in environments with minimal building coverage.

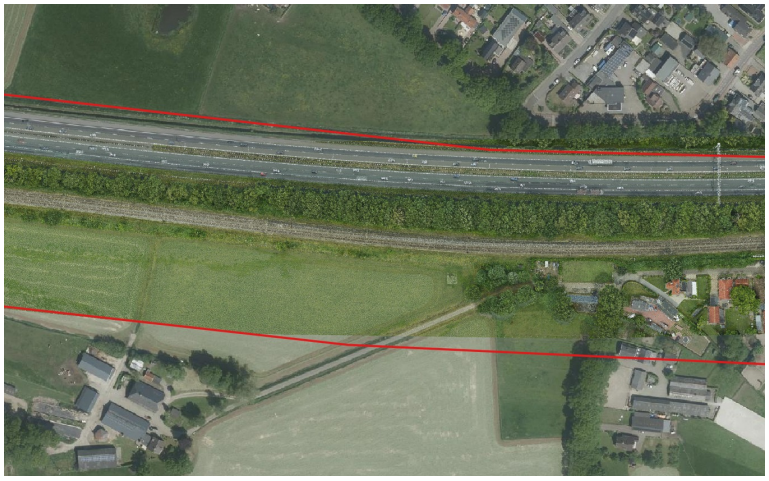


Figure 3.2: Isolated Prorail-Spoorinbeeld point cloud data, red lines show the boundary of the data

3.2 Data Description

This research uses multi-source point cloud datasets acquired from different platforms and time periods. The datasets originate from three different sources. The first is railway point cloud data provided by ProRail through the SpoorInBeeld project as open-access data. The second is the national elevation dataset of the Netherlands, provided by the AHN project, also as open-access data. The third is highway point cloud data provided exclusively for this research by Rijkswaterstaat.

3.2.1 ProRail's SpoorInBeeld

ProRail is the organisation responsible for managing and maintaining the Dutch railway network. Its core responsibilities include ensuring the safe and timely transport of passengers and goods, maintaining railway infrastructure, and coordinating with rail carriers and contractors ProRail 2024. As part of its efforts to support infrastructure monitoring and planning, ProRail provides access to spatial data through the SpoorInBeeld platform ProRail SpoorInBeeld 2024. This platform offers a variety of visual and geospatial datasets related to the railway and its surroundings, including aerial imagery, panoramic photographs, and laser point clouds, which are valuable for analysis, visualisation, and innovation in railway management.

Among the available datasets, the laser point clouds from the SpoorInBeeld programme cover large parts of the Dutch railway network and nearby areas. These data were collected using LiDAR systems mounted on helicopters, trains, and ground equipment between 2018 and 2022. The point clouds have been checked for both positional accuracy and the availability of intensity information. The positional accuracy is less than 12 cm absolute and less than 6 cm relative, based on the Dutch national coordinate system (RD/NAP) Heide et al. 2024. This dataset is listed in the national overview of point clouds created as part of the Integral Height Facility Netherlands (IHN) project, where it is included as one of the point cloud sources being studied and integrated Heide et al. 2024. As such, the ProRail dataset is relevant not only for railway maintenance but also for research on the alignment of multi-source point cloud data.

3.2.2 Actueel Hoogtebestand Nederland (AHN)

The Actueel Hoogtebestand Nederland (AHN) is a national programme that provides detailed and accurate elevation data for the entire Netherlands. It is a collaboration between water boards, provinces, and Rijkswaterstaat, aimed at supporting water system and defence management AHN 2024. AHN data are collected using airborne laser scanning and are made publicly available as open data. To facilitate access to these datasets, AHN offers a platform known as the AHN Data Room, where users can download various versions of the dataset, including ground and surface models in raster format and raw point clouds in LAZ format AHN Dataroom 2024. Given the large volume of data, the GeoTiles platform was developed to improve accessibility and usability by dividing the dataset into smaller, manageable tiles. GeoTiles provides a consistent tiling structure and map-based access to AHN products, allowing researchers and practitioners to efficiently view and download data relevant to their area of interest GeoTiles.nl 2024.

The AHN dataset is available in multiple versions, including AHN2 through AHN5, each collected in phases across defined geographical plots. The point clouds are referenced to the national RD/NAP coordinate system and are typically delivered in LAZ format. Depending on the version, the point density ranges from approximately 6 to over 10 points per square metre, with vertical accuracy requirements improving over time, up to 3 cm stochastic error in AHN5 Heide et al. 2024. In addition to being a widely used national elevation source, AHN is included in the national inventory of point clouds compiled under the Integral Height Facility Netherlands (IHN) project, which supports harmonisation of heterogeneous point cloud data across the country Heide et al. 2024. Given its high accuracy and nationwide coverage, AHN serves as a valuable reference dataset in this research for evaluating and aligning LiDAR data from other sources.

3.2.3 Rijkswaterstaat Point Clouds Datasets

Rijkswaterstaat is the executive agency of the Dutch Ministry of Infrastructure and Water Management. It is responsible for the design, construction, management, and maintenance of the Netherlands' main infrastructure networks, including national roads, waterways, and water systems. The agency plays a key role in ensuring safe and efficient mobility, managing water levels, and protecting the country against flooding Rijkswaterstaat n.d. Through its work, Rijkswaterstaat supports a sustainable, accessible, and climate-resilient living environment across the Netherlands.

Rijkswaterstaat actively uses point cloud data to support a variety of infrastructure-related tasks, including the creation of DTM, drive-through visualisations, tunnel modelling, and the maintenance of the digital topographical file (Digitaal Topografisch Bestand). These datasets are typically acquired through project-based data collection, particularly for large-scale roadworks, and are subject to different quality requirements depending on the intended application and environmental context (e.g. dry or wet conditions). Required accuracies generally range from 2 to 15 centimetres, depending on the object type and location Heide et al. 2024. Point cloud data is an important part of Rijkswaterstaat's 3D geodata system, which helps ensure consistent use of spatial data across the organisation. For this reason, point clouds collected or managed by Rijkswaterstaat are also listed in the national point cloud inventory under the Integral Height Facility Netherlands (IHN) project.

3.2.4 Study Area

In Section 3.1, a case is proposed to simulate the condition where road marking is used as an alternative benchmark to check the alignment and to co-register the railway point cloud with AHN point cloud data. To better align with the proposed case, this study focuses on a location south of the village of Terschuur, in the municipality of Barneveld, province of Gelderland.

The area was chosen due to the availability of high-resolution point cloud data collected by ProRail as part of the SpoorInBeeld project, which typically covers a corridor extending approximately 50 metres on either side of the railway (see yellow line in Figure 3.5). In this specific location, a highway runs parallel to the railway and lies entirely within the buffer zone of the dataset (see red line in Figure 3.5). As a result, a substantial number of road markings on the highway are captured in the point cloud. These road markings

provide a consistent and well-distributed set of features that can potentially be used to assess alignment accuracy and co-register the railway dataset with other point clouds, particularly in areas where the presence of buildings is limited.

Figure 3.3: Area of interest

The ProRail SpoorInBeeld point cloud data were downloaded from ProRail SpoorInBeeld 2024 in March 2025. Several datasets acquired using different platforms are available via the portal, including helicopter-, vehicle-, and tripod-based LiDAR, collected across multiple epochs. However, at the time of download and area selection, only point cloud data acquired by helicopter in the years 2019, 2021, and 2023 were available. These datasets include intensity information, which is essential for the extraction of road markings. In terms of classification, each dataset contains only two classes: railway and non-railway.

AHN point cloud data are used as reference data to assess alignment and to do the co-registration. The AHN versions used in this research are AHN3, AHN4, and AHN5, which were downloaded per AHN sub-grid via the GeoTiles platform and downloaded in March 2025. To identify which grids needed to be downloaded, an Area Of Interest (AOI) polygon, digitised from the boundary of the downloaded ProRail point cloud, was overlaid with the AHN grid layout to determine which grids the AOI intersected. This resulted in the selection of grids 32EZ1_19 and 32EZ1_20, which were subsequently downloaded and clipped using the AOI polygon. The AHN point clouds include intensity values, allowing them to be used for automatic road marking extraction. Unlike the ProRail datasets, the AHN data are classified into six categories: ground, building, works of art, water, high voltage, and other.

All point clouds used in this research are provided in the Dutch national coordinate reference system (RD New / EPSG:28992). While they share a common spatial reference, the datasets vary in both intensity value ranges and point density. The intensity range of each dataset was examined in CloudCompare using the scalar field visualisation of intensity values. To assess point density, the local surface density was computed in CloudCompare using the "Compute Geometric Features" tool, with a spherical neighbourhood radius set to one metre. This calculation provides an estimate of the average number of points per square metre for each point cloud. The resulting intensity ranges and average density values for all datasets used in this study are summarised in Table 3.1.

Table 3.1: Intensity range and point density of the datasets used in this research

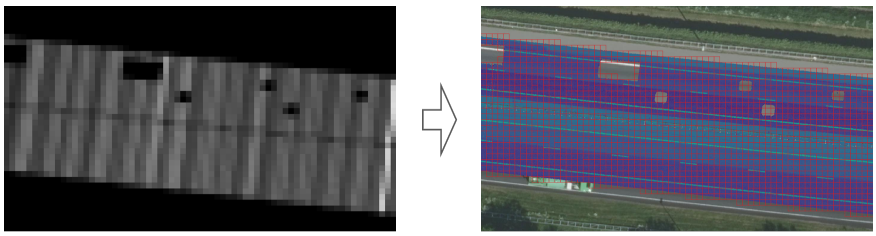
Dataset	Year	Mean Density (points/m²)		Intensity Range
		Inside AOI	Road Only	
ProRail SpoorInBeeld	2019	66	70	45,000 ∼ 55,000
ProRail SpoorInBeeld	2021	134	144	25,000 ∼ 40,000
ProRail SpoorInBeeld	2023	155	170	27,000 ∼ 36,000
AHN3	2018	10	10	110 ∼ 380
AHN4	2020-2022	23	24	1100 ∼ 1,500
AHN5	2023	17	17	1230 ∼ 4,400

The density values and intensity ranges listed in Table 3.1 were calculated specifically for the point cloud data within the defined AOI used in this study. The density represents the average number of points per square metre within the AOI, rather than across the entire dataset extent. Similarly, the reported intensity range corresponds to the minimum and maximum intensity values of road markings observed in the AOI. This local calculation makes sure that the values shown in the table match the actual point cloud data used for road marking extraction and alignment in this study.

Calculation of Mean Density

There are two types of density values presented in Table 3.1. The values in the "Inside AOI" column represent the density calculated from all points within the research AOI, including all classes available in the point cloud data. In contrast, the "Road Only" density refers to the density calculated exclusively from points representing the road surface, specifically those classified as ground, where the road markings are located. This differentiation ensures that the density values more accurately reflect the environment relevant to the road markings

To calculate the point cloud density, density rasters were first generated using the "Point Cloud Extraction - Density" tool from the PDAL library in QGIS. After creating the density rasters, they were converted into vector polygons using the "Polygonize" tool in QGIS, where each polygon represents a cell with a corresponding density value. The mean density was then computed by averaging the values of all polygons.

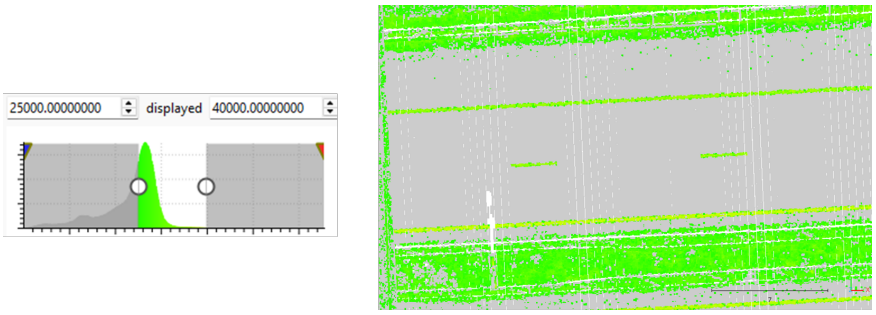


(a) (b)

Figure 3.4: Illustration of density calculation. (a) Density raster of the point cloud. (b) Polygonized density raster.

Determination of Intensity Range

The intensity ranges shown in Table 3.1 represent the range of road marking intensities that were manually filtered using the intensity display settings in CloudCompare. This selection was performed manually by adjusting the minimum and maximum intensity values based on the visual appearance of the point cloud. As a result, the selected ranges are subjective. However, they are useful for illustrating the variation in intensity range values across the different point cloud datasets used in this research



(a) (b)

Figure 3.5: Illustration of intensity range determination. (a) Manual adjustment of intensity range. (b) Displayed point cloud.

3.3 Automatic Extraction of Road Markings

In this research, road marking data were provided by Rijkswaterstaat, who performed the extraction using the method proposed in Heide 2024. This method is designed to work with LiDAR point cloud that contains intensity information, regardless of whether it was acquired through airborne, mobile, or terrestrial laser scanning. It applies an adaptive intensity threshold approach that adjusts locally to differences in intensity range across the dataset. This is particularly useful to extract benchmarks for co-registration, where overlapping point clouds may come from different sensors or acquisition settings and therefore have varying intensity ranges. By using intensity as the main input feature, the method is able to detect linear road marking features consistently without relying on RGB imagery or converting the point cloud into raster format.

Since the extraction process was conducted externally and lies outside the scope of this study, only the resulting road marking features were used as input for the co-registration workflow. A brief summary of the method is included in the related work Section 2.2 for reference. However, this research contributes to finding the threshold values of geometric properties used to describe road marking clusters. In particular, the method proposed in Heide 2024 applies minimum and maximum threshold values to shape descriptors, derived from the eigenvalues of the local neighbourhood, such as linearity, planarity, sphericity, anisotropy, the sum of eigenvalues, and change in curvature (see Table 2.4). These thresholds are used to filter out clusters that do not have the typical geometric characteristics of road markings, helping to distinguish them from other high-intensity objects in the point cloud.

The workflow to compute the road marking clusters geometry features threshold is shown in Figure 3.6

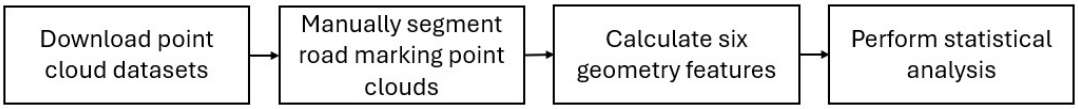


Figure 3.6: Workflow to compute road marking clusters geometry features threshold

To calculate the threshold values of the road marking clusters' geometric features, approximately 100 road markings were manually segmented from the AHN3, AHN4, and AHN5 datasets, covering various locations across the Netherlands. The AHN datasets were chosen as the baseline for calculating the geometric feature thresholds of road marking clusters because they serve as the reference datasets in the co-registration process. Since other point cloud datasets (e.g., from SpoorInBeeld) are aligned to AHN, it is important that the definition of road marking geometry is consistent with the characteristics present in AHN data.

Given the large number of road markings available throughout the Netherlands, selecting samples for segmentation can be challenging. To ensure a representative and evenly distributed sample, approximately 60 points were randomly placed across the country as starting locations, as illustrated in Figure 3.7. Around each of these points, one to three road markings from either the AHN3, AHN4, or AHN5 datasets were selected for manual segmentation.

Figure 3.7: Locations used to choose road marking examples

The sampling locations were distributed across the entire Netherlands to ensure that the selected road markings represent a wide range of geographic, environmental, and acquisition conditions. Since AHN datasets are collected over multiple years and regions using different flight paths and sensors, road markings may vary in appearance, intensity, and point density. By distributing the sample locations across the country and incorporating three different AHN versions (AHN3, AHN4, and AHN5), the resulting threshold values for geometric features are expected to be more generalisable and robust across varying regions and dataset versions.

After the random points were distributed, it was necessary to identify which sub-grids from GeoTiles.nl 2024 each point falls into, in order to simplify the navigation and download of the relevant data. To achieve this, the sub-grid shapefile data from GeoTiles.nl 2024 is downloaded and overlaid with the sample points as illustrated in Figure 3.8, making it easier to determine the correct sub-grid names for downloading the corresponding point cloud datasets.

Figure 3.8: GeoTiles.nl grid showing the road marking sample locations used to select the point cloud data

The road markings to be segmented were randomly selected around a set of randomly placed points and extracted from the AHN3, AHN4, and AHN5 datasets. The segmentation was done manually in CloudCompare using the "Segment" tool, and all segmented files were saved in a single folder to allow automatic processing using a Python script. These segments were used as input to calculate six geometry features for each road marking cluster.

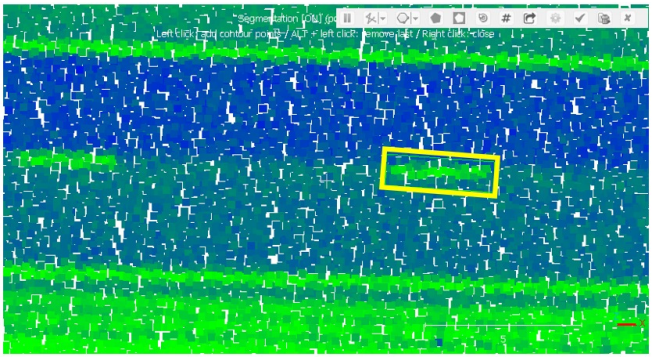


Figure 3.9: Illustration of road marking segmentation in CloudCompare, yellow polygon used to select the points inside the polygon then segment it as a new dataset

As illustrated in Figure 3.9, the segmentation process is performed by manually drawing a polygon around the points, using the intensity values for visual guidance. The polygon is drawn to enclose points that indicate high intensity, represented in this case by bright green colouring. However, during the polygon drawing process, some low-intensity points, possibly corresponding to pavement surfaces, may still be included within the segment. These points could influence the computed geometry feature values of the cluster.

To further refine the segmented road marking clusters and reduce the influence of unwanted points, such as low-intensity pavement points accidentally included during manual segmentation, a z-score filtering method was applied to the intensity values of each segmented point cloud. This statistical approach, commonly used in outlier detection (Benallal et al. 2022), calculates how far each intensity value deviates from the mean in terms of standard deviation, allowing the identification and removal of non-representative points. In this context, points with unusually low intensity are excluded, while high-intensity points, which are more typical of road marking materials, are retained.

To achieve this, each intensity value x is converted into a standardised z-score using the following formula:

$$z = \frac{x - \mu}{\sigma} \tag{3.1}$$

where μ is the mean and σ is the standard deviation of all intensity values within the segment.

Points are retained for further processing if they satisfy the following condition:

$$z \geq z_{\text{threshold}} \tag{3.2}$$

In this study, the value of $z_{\text{threshold}}$ was determined by manually testing several threshold values on a sample road marking segment. A value of -1.5 was selected based on tests conducted on sample segment "RM_21", as presented in Section 5.2. This relatively low threshold allows a larger number of points to be retained, including those with moderately

high intensity. The reason for retaining a larger portion of points in each segment is that, during the segmentation process, the polygon used to segment the road markings was drawn as small as possible to keep only high-intensity points, with the assumption that most points inside the polygon truly belong to the road marking. As a result, using a very low z-score threshold helps to remove only the few points with very low intensity that were accidentally included.

The 3D coordinates of these filtered points are then used to calculate a covariance matrix, which shows how the points are spread in space. From this matrix, eigenvalue decomposition is applied using the NumPy package (Harris et al. 2020) to obtain three eigenvalues, which represent how the points vary along different directions. These eigenvalues are then used to compute six geometry features: linearity, planarity, sphericity, anisotropy, sum of eigenvalues, and curvature, as shown in Table 2.4. All feature values from the samples are collected for further analysis.

After the geometry features are calculated, statistical filtering is applied to determine reliable threshold values for each feature. This is done using the Interquartile Range (IQR) filtering method, which removes outliers by keeping only the values that fall within a reasonable range around the central portion of the data (Dekking et al. 2005).

The IQR is defined as:

$$\text{IQR} = Q_3 - Q_1 \tag{3.3}$$

where Q_1 and Q_3 are the first and third quartiles, respectively.

Values are retained if they fall within the following range:

$$[Q_1 - 1.5 \cdot \text{IQR},\ Q_3 + 1.5 \cdot \text{IQR}] \tag{3.4}$$

The filtered feature values are then used to determine the minimum and maximum range for each geometry feature. Table 3.2 presents the minimum and maximum values of each geometric feature after the IQR filtering. These ranges are considered the adjusted thresholds and are intended to better represent typical road marking characteristics. They can later be used to help distinguish road marking clusters from other objects in different point cloud datasets during the co-registration process.

Table 3.2: Minimum and maximum values of each geometric feature after IQR filtering

Name	**Min Value**	**Max Value**
Linearity	0.9265	0.9990
Planarity	0.0009	0.0732
Sphericity	0.0000	0.0011
Anisotropy	0.9989	1.0000
Sum of eigenvalues	0.0362	1.2091
Change in curvature	0.0000	0.0011

3.4 Manual Digitisation of Road Markings

To evaluate the quality of automatically extracted road markings, a reliable reference dataset is required. The reference road markings data were created through manual digitisation of road markings directly from the point cloud datasets. This process involves visually identifying the road marking features, using their higher intensity values relative to the surrounding points as a cue. The manually digitised road markings serve as a benchmark for assessing the quality of the road marking extraction. By comparing the extracted data to this reference, it becomes possible to quantify the performance of the extraction method and evaluate the accuracy of the extracted road marking model, thereby providing insights into how variations in road marking quality may affect subsequent co-registration accuracy. To achieve a comparable intensity value range when displaying road marking points, in line with the automatic extraction method proposed in Heide 2024, the road points are first segmented, followed by ground filtering to remove noise. The workflow followed for reference road markings creation is illustrated in Figure 3.10.

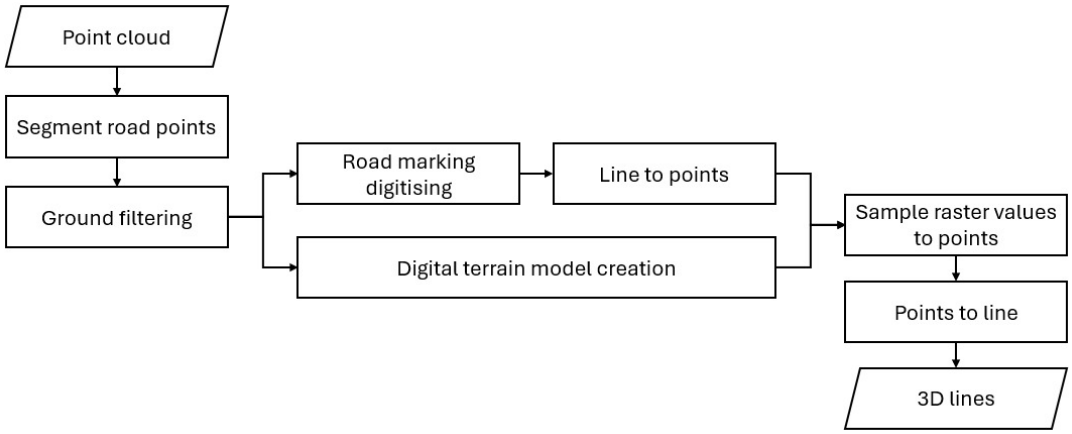


Figure 3.10: Workflow to generate reference road markings

The first step involves segmenting the road points from the rest of the point cloud. This is done to ensure that the intensity value range used for visualisation includes only the points within the road corridor where road markings are expected to be located. By limiting the data to road areas, objects outside the corridor that have high-intensity values, such as the unidentifiable object shown in Figure 3.11, can be excluded. These non-road objects may otherwise interfere with the road marking digitisation process.

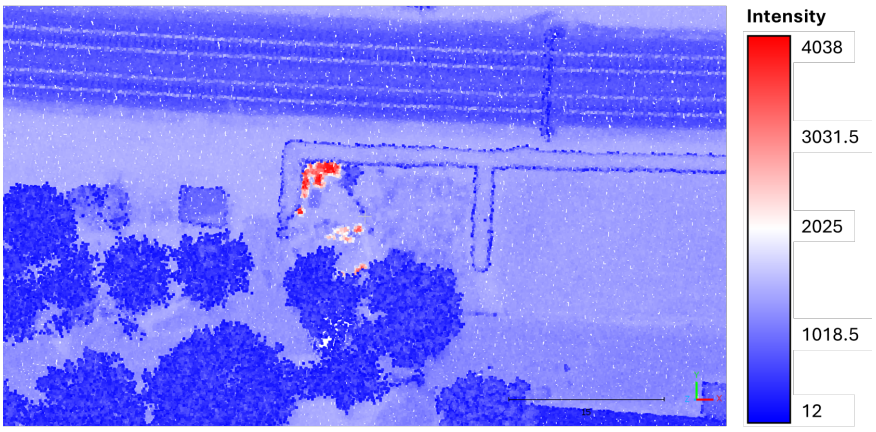


Figure 3.11: Unidentified object with high intensity value outside road corridor

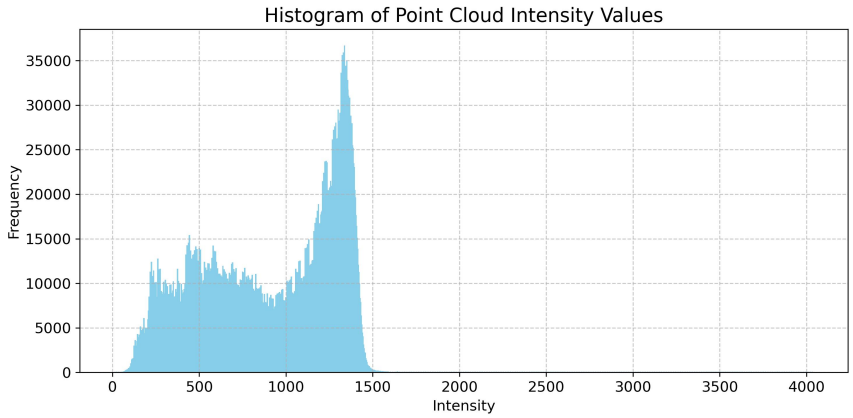


Figure 3.12: Histogram of intensity values of point cloud dataset in Figure 3.11

The next step involves removing noise and objects other than the road pavement from the point cloud. While the AHN data already includes classification for cars and noise on the road, the ProRail point cloud only distinguishes between railway and non-railway points. As a result, vehicles on the road are still present in the dataset, as illustrated in Figure 3.13. Therefore, ground classification must be performed to isolate the road surface. The ground classification to remove unwanted objects on the road corridor is done in CloudCompare using Cloth Simulation Filter (CSF) plugin by Zhang et al. 2016.

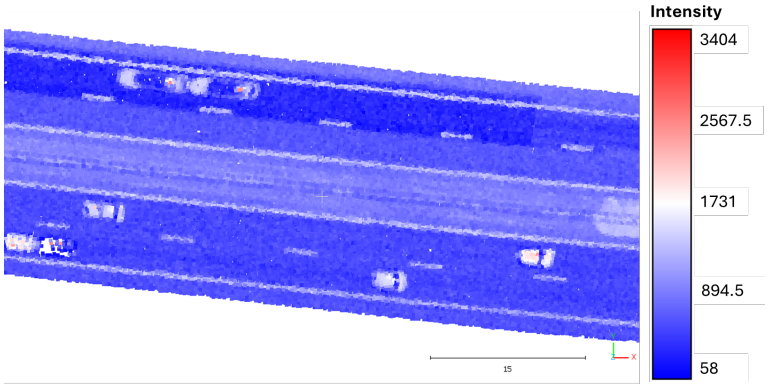


Figure 3.13: Detected cars on the road corridor

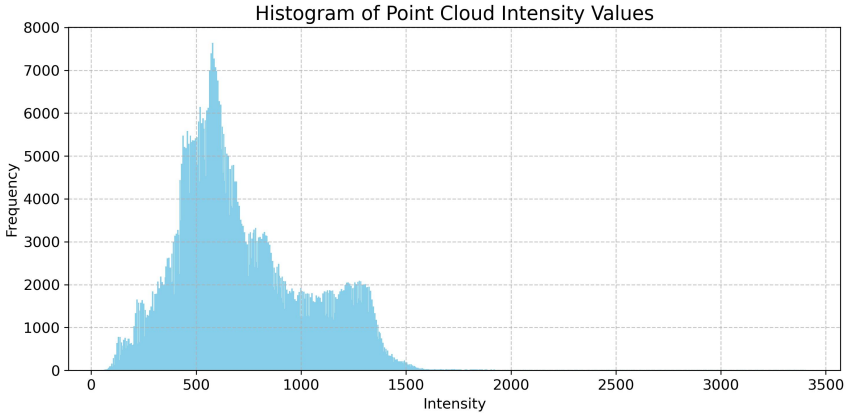


Figure 3.14: Histogram of intensity values of point cloud dataset in Figure 3.13

After the road points have been segmented and ground classified, it is assumed that the intensity range in the point cloud used for manual digitisation is comparable to that used in the automatic road marking extraction. The process allows the manual digitisation to closely reflect the conditions under which the automatic method operates, making it suitable as a reference for evaluation.

Since the digitisation of road markings is performed using a 2D view, additional processing is required to assign height information to each digitised line. To address this, the previously ground-filtered points are rasterised into a DTM, which is then overlaid with the digitised road markings to transfer elevation values from the DTM to each line. The rasterisation of the ground-classified road point cloud is carried out in CloudCompare using the ”Rasterize” tool. The result is shown in Figure 3.15

Figure 3.15: Generated digital terrain model

To assign elevation values from the DTM to the digitised 2D road marking lines, each line is first simplified by converting it into two points representing its start and end positions. These two points are then used to sample elevation values from the DTM raster, where each point retrieves the ground height from the corresponding raster pixel beneath it. The sampled elevation values are then assigned to the start and end points of the line, allowing the reconstruction of a 3D line with height and slope information. This results in 3D lines that are comparable to the extracted 3D lines produced by the automatic road marking extraction.

3.5 Extraction Quality Measurement

To evaluate the quality of the extracted road markings, a comparison is made between the automatically extracted features and the manually digitised road markings. This evaluation consists of two parts: detection accuracy metrics and geometric comparisons. The first part assesses whether the right road markings are found, missed, or wrongly included. The second part evaluates how precisely the shape and position of the automatically extracted road markings match the manual digitised road markings. Together, these assessments help determine whether the extracted features are suitable for use as benchmarks in point cloud co-registration.

3.5.1 Detection Accuracy

Detection accuracy metrics focus on how well the algorithm performs in identifying road markings. The use of these metrics in this research is inspired by established practices in automatic feature extraction, where similar metrics have been applied to evaluate road extraction performance from aerial imagery Heipke et al. 1997. These metrics indicate how many features were correctly detected, how many were missed, and how many were incorrectly extracted. Rather than serving as an absolute measure of accuracy, they are used to compare the performance of the road marking extraction method across different datasets. This relative evaluation helps assess whether the method is robust and consistent when applied to point clouds from varying sources or acquisition conditions. Additionally, applying these metrics makes it possible to explore whether the completeness, correctness, and overall quality of the extraction method influence the accuracy of the co-registration process.

This assessment is based on three common outcomes: **true positives (TP)**, **false positives (FP)**, and **false negatives (FN)**. A true positive (TP) refers to a road marking that is correctly

detected by the algorithm and also exists in the reference road markings. A false positive (FP) is a feature that is incorrectly identified as a road marking but is not present in the reference road markings, for example, duplicated detections of the same marking or other high-intensity objects mistakenly classified as road markings. A false negative (FN) occurs when a true road marking is present in the reference road markings but is not detected by the extraction method.

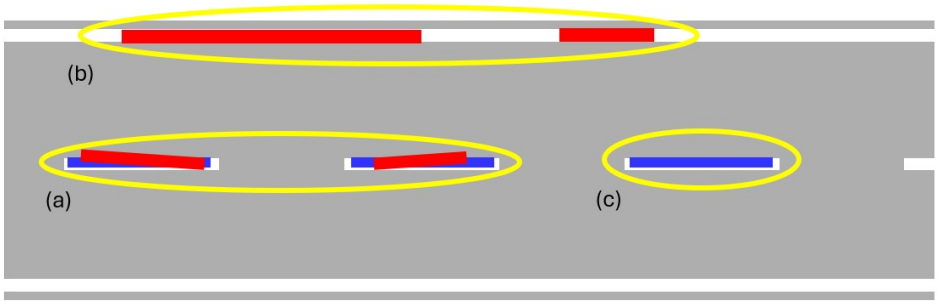


Figure 3.16: Road markings detection cases. Blue lines represent manually digitised road markings, red lines represent automatically extracted road markings: (a) True positive (TP). (b) False positive (FP). (c) False Negative (FN).

To quantify these outcomes, the comparison is based on the spatial proximity of road marking centroids. First, the centroid of each road marking is computed for both the manually digitised and the automatically extracted road markings. A nearest neighbour search is then performed from each manually digitised road marking's centroid to the centroids of the automatically extracted road markings, with a search radius. If a corresponding extracted road marking is found within the range, it is considered a true positive (TP). Any extracted road marking that has already been matched to a manually digitised road mark is marked as part of a correct detection. Extracted segments that do not correspond to any manually digitised road mark (i.e., not paired in the matching process) are considered false positives (FP). On the other hand, if a manually digitised centroid does not have any nearby extracted counterpart within the defined range, it is counted as a false negative (FN). Cases of road marking detection are illustrated in Figure 3.16

Based on these categories, three evaluation metrics are calculated: completeness, correctness, and quality. Completeness measures how many of the actual road markings in the manually digitised road markings were successfully identified. It is calculated as:

$$\text{Completeness} = \frac{TP}{TP + FN} \tag{3.5}$$

Correctness shows the proportion of extracted road markings that are actually correct. It is calculated as:

$$\text{Correctness} = \frac{TP}{TP + FP} \tag{3.6}$$

Quality provides an overall score by combining both missed detections and incorrect extractions. It is defined as:

$$\text{Quality} = \frac{TP}{TP + FP + FN} \tag{3.7}$$

These three metrics are used to compare how well the road marking extraction method works with different types of input data, such as changes in point density or intensity range. The goal is to see if the method gives consistent results under different conditions. The extraction accuracy values are then used in the final evaluation to help explain any differences in the alignment results.

3.5.2 Line Extraction Accuracy

While the detection accuracy metrics show whether the correct road markings are detected, they do not measure how accurate the extracted features are in terms of shape, position, or orientation. Since the road markings are extracted as line features for co-registration, their spatial accuracy may also affect the final alignment. To investigate this, an evaluation is carried out to measure how accurately the road markings are modelled and to explore whether their spatial accuracy plays a role in influencing the accuracy of the co-registration process. To address this, a geometric comparison is carried out to assess how closely the extracted road markings match the manually extracted road markings.

Three aspects are considered as parameters for measuring the accuracy of the extracted line features: length, centroid position, and orientation. The length error measures the difference in segment length between the extracted and reference lines. The position error is calculated by comparing the 3D coordinates of the midpoints of each segment, providing insight into any spatial shifts. For orientation, the accuracy is evaluated in two components: the horizontal angle error, which captures the difference in direction within the x–y plane, and the vertical angle error, which reflects slope differences based on elevation values derived from the DTM. The illustration of these aspects is shown in Figure 3.17

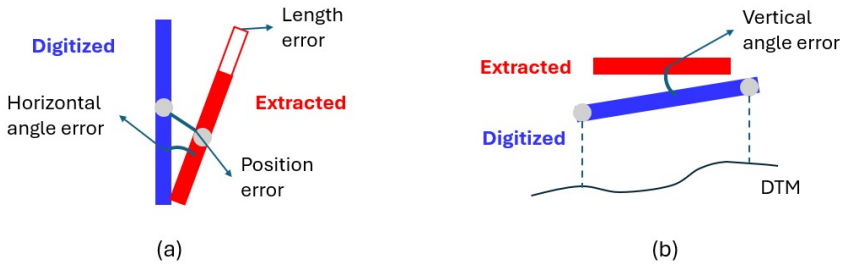


Figure 3.17: Aspects for evaluating line accuracy. Blue lines represented manually digitised lines, red lines represented automatically extracted lines. (a) Top view. (b) Side view

3.6 Correspondence Establishment

After evaluating the quality and geometric accuracy of the extracted road markings, the next step is to establish correspondences between features in different point cloud datasets. In this research, road markings are used as the benchmark to align datasets captured from different sources or at different times. To perform co-registration, it is necessary to identify matching pairs of road markings between the reference and target datasets.

The process begins by calculating the 3D centroid of each road marking segment in both datasets. Each road marking is modelled as a 3D line segment with two endpoints. The centroid of a line is calculated as the average of its start and end point coordinates:

$$C = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}, \frac{z_1 + z_2}{2} \right) \tag{3.8}$$

where (x_1, y_1, z_1) and (x_2, y_2, z_2) are the coordinates of the start and end points of the line segment, and C is the centroid of the line.

A nearest neighbour search is then performed to identify potential correspondences, where each road marking centroid in the reference dataset is matched to its closest road marking centroid in the target dataset, using a defined maximum search distance to avoid false matches. This distance threshold is chosen based on how far apart the same road markings might appear in the two datasets before they are aligned. The illustration of correspondence finding shown in Figure 3.20

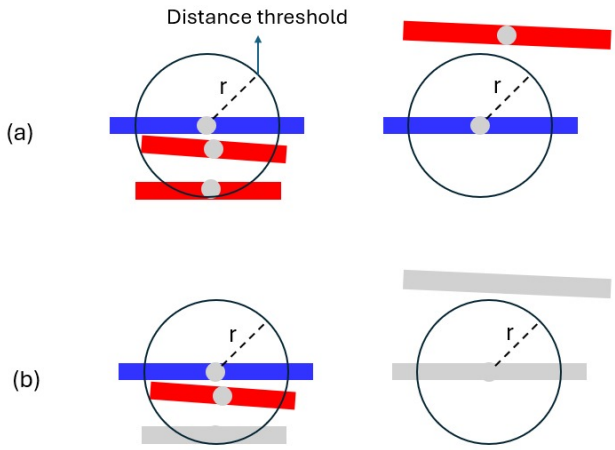


Figure 3.18: Correspondence finding. Blue lines are reference road markings, red lines are target road markings, grey lines are removed road markings. (a) Before calculation. (b) After calculation.

The distance threshold in the nearest neighbour search is used to filter an initial set of candidate correspondences based on spatial proximity between road marking centroids. At this stage, the goal is not to identify precise matches, but to include all possible correspondences that could later be verified through later filtering step. Therefore, the threshold should be

wide enough to account for expected positional differences between datasets, while still excluding clearly unrelated features.

Only pairs that fall within this distance threshold are retained as candidate correspondences. These pairs are later used as input in the RANSAC-based alignment process, where their geometric consistency is evaluated across multiple transformation estimations. In each RANSAC iteration, a small subset of candidate pairs is randomly selected to compute a rigid transformation.

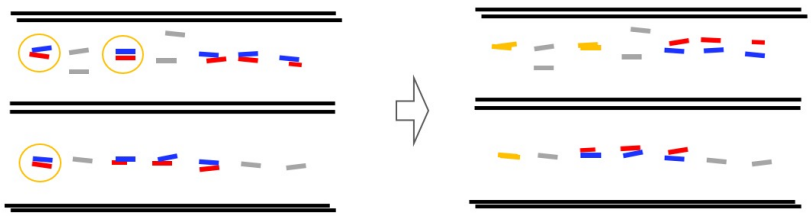


Figure 3.19: Illustration of alignment during a RANSAC-based iteration. Blue lines represent road markings in the reference dataset, red lines represent road markings in the target dataset, and yellow-circled pairs are the subset of correspondences used to estimate the transformation between the two datasets

In each RANSAC iteration, after estimating the transformation using a randomly selected subset of correspondence pairs, the remaining candidate correspondences are evaluated to determine whether they qualify as inliers. This is done by applying the estimated transformation to the target dataset and then measuring the distance between each transformed target centroid and its corresponding reference centroid. If the centroid-to-centroid distance falls within a predefined tolerance, the pair is considered an inlier for that iteration. This threshold is set smaller than the initial nearest neighbour threshold to further filter the candidate correspondences and identify those that align closely after transformation.

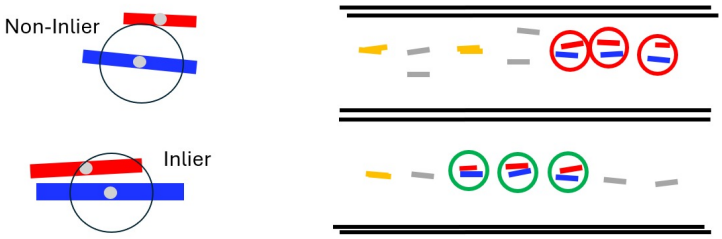


Figure 3.20: Illustration of RANSAC inliers evaluation. Blue lines represent road markings in the reference dataset, red lines represent road markings in the target dataset, and green-circled pairs are counted as inliers, red-circled pairs are counted as non-inliers

In this research, the distance threshold used to determine whether a correspondence is considered an inlier is set to 11.2 centimetres. This value is based on the official accuracy

specifications published in the IHN's work package 1 report (Heide et al. 2024), which mentioned the expected measurement accuracy for different types of features in Dutch point cloud datasets. Road markings fall under the category of hard topography, which includes clearly defined objects such as road lines. For this category, the specified absolute accuracy is 5 cm in the horizontal direction and 10 cm in the vertical direction. These values are combined into a single 3D positional tolerance using the root sum of squares:

$$\textit{inlier distance threshold} = \sqrt{5^2 + 10^2} = 11.2 \text{ cm} \tag{3.9}$$

This threshold is used during the RANSAC process to check whether a transformed pair of road markings can be considered an inlier. However, this value is based on the expected accuracy of a single dataset, while the RANSAC method compares features from two different datasets. As a result, the actual alignment error could be larger. Even though it's like that, this threshold provides a consistent and practical reference derived from national standards, and is applied here to help filter out unreliable correspondences.

The process of inliers counting are illustrated in Figure 3.21. In this example, the inlier count is treated as an attribute associated with each road marking. Whenever a pair is classified as an inlier in a given iteration, it receives a plus 1 increment in its inlier count attribute. As shown in the figure, RM7, RM8, and RM9 are the only pairs that were considered inliers during the current iteration, each receiving a plus 1 update. The rest of the pairs remain unchanged, with an inlier count of plus 0, meaning they did not support the transformation in that iteration.

Iteration-n		
ID	Inlier_count	Total iteration
RM 1	+0	1000
RM 2	+0	1000
RM 3	+0	1000
RM 4	+0	1000
RM 5	+0	1000
RM 6	+0	1000
RM 7	+1	1000
RM 8	+1	1000
RM 9	+1	1000

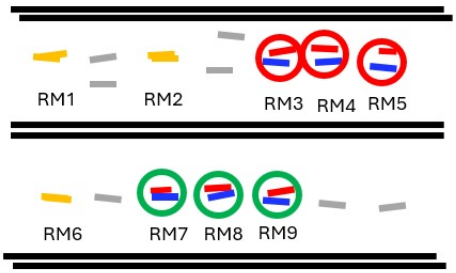


Figure 3.21: Example of inlier counting in a single RANSAC iteration

Figure 3.22 shows the example of result attribute table after completing 1000 RANSAC iterations. This table reflects the final accumulated inlier count values for each road marking. These values indicate how consistently each correspondence supported the estimated transformations throughout the alignment process. Pairs such as RM1, RM3, RM4, RM6, RM8, and RM9 have high inlier counts, suggesting that they are geometrically stable and reliable. In contrast, pairs like RM2, RM5, and RM7 have low inlier counts and are considered less stable, as they rarely contributed to valid transformations. This information can be used

to identify and filter out unstable correspondences before estimating the final transformation.

Iteration-1000		
ID	Inlier_count	Total iteration
RM 1	998	1000
RM 2	122	1000
RM 3	835	1000
RM 4	923	1000
RM 5	32	1000
RM 6	890	1000
RM 7	14	1000
RM 8	841	1000
RM 9	915	1000

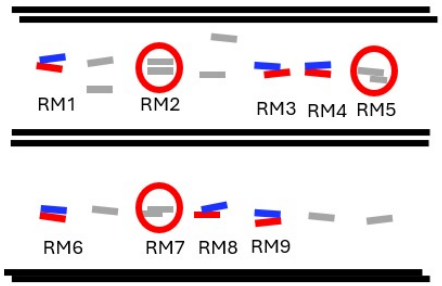


Figure 3.22: Example of accumulated inlier counts after 1000 RANSAC iterations. Red-circled pairs are pairs that considered to be unstable

A pair of road markings is considered geometrically stable if it aligns well with the transformation in many RANSAC iterations. A high inlier count means the pair frequently fits the estimated transformation, and it is assumed that the road markings constructing the pair are well-extracted. On the other hand, a low inlier count may indicate small errors in the feature, such as a 3D centroid error. The approach is used to help filter out road markings with low extraction quality. The underlying assumption is that if a road marking is not accurately extracted, either in the reference or target point cloud, it will not align well during the transformation in a RANSAC iteration and will therefore be less likely to be counted as an inlier.

However, the inlier count is only an indirect way to assess feature quality. A high count does not always mean that the features in the pair have been accurately extracted. It could also occur if both features have similar centroid shifts in magnitude and direction. Although like that, the inlier count remains useful for identifying correspondences that are less reliable and may be affected by extraction errors that impact the alignment.

3.7 Rotation and Translation Values Calculation

Once the correspondences between road marking line features have been established, the next step is to compute the transformation matrix required to align the target dataset to the reference. This transformation consists of a rigid 3D motion, including both rotation and translation, and is estimated based on the geometric differences between the matched line segments.

This research applies a RANSAC-weighted centroid alignment method, which extends the least-squares approach introduced in Umeyama 1991. In this method, each road marking is represented by its 3D centroid, and a weight is assigned to each matched pair based on its inlier count recorded during the correspondence establishment phase. The inlier count

for each pair is normalised by the total number of RANSAC iterations to produce a relative weight:

$$w_i = \frac{c_i}{T} \tag{3.10}$$

where c_i is the inlier count for the i-th pair, and T is the total number of RANSAC iterations. This produces a relative weight $w_i \in [0, 1]$, which reflects how consistently a pair supports the alignment transformation. Pairs with higher weights have more influence in the calculation of the transformation matrix, while those with low weights contribute less.

With these weights defined, the transformation calculation proceeds by incorporating them into the standard least-squares. Instead of using simple averages to compute the centroids, as done in the original formulation shown in Equation 2.2, the centroids of the matched pairs are now calculated using a weighted average. This means that pairs with higher weights have more influence on the final centroid values.

The weighted centroids are computed as:

$$\bar{\mathbf{p}} = \sum_{i=1}^{N} w_i \mathbf{p}_i, \quad \bar{\mathbf{q}} = \sum_{i=1}^{N} w_i \mathbf{q}_i \tag{3.11}$$

Following the computation of the weighted centroids, the next step is to calculate the weighted cross-covariance matrix. This matrix describes how the positions of the reference and target centroids relate to each other after they've both been shifted so their average position is at the centre. In the weighted version, each pair contributes to the covariance matrix according to its assigned weight, so that pairs with higher weights influence the estimated transformation more than those with lower weights. This step is a modified version of the original cross-covariance calculation in Equation 2.4, where the formula is adjusted so that each matched pair contributes based on its assigned weight.

The weighted cross-covariance matrix $\mathbf{H}_w$ is:

$$\mathbf{H}_w = \sum_{i=1}^{N} w_i (\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{q}_i - \bar{\mathbf{q}})^{\mathrm{T}} \tag{3.12}$$

In this formula, the weight w_i controls how much each matched pair affects the cross-covariance matrix $\mathbf{H}_w$. Pairs with higher weights, those that were often identified as inliers during the RANSAC process, contribute more to the matrix. These pairs are assumed to be more reliable, so the transformation is influenced more by them than by less consistent pairs. This helps the alignment focus on the stable geometric structure between the datasets and reduces the impact of noisy or incorrect matches.

Following the same SVD process in Equation 2.5, the optimal weighted rotation matrix is then obtained using Equation 2.6 and Equation 2.7. The translation vector is then also obtained using the Equation 2.8 after optimal rotation matrix is calculated.

By weighting each matched pair based on its inlier count from the correspondence establishment process, the method gives more influence to geometrically stable pairs during the transformation calculation. This helps reduce the impact of noisy or unreliable matches, which may result from extraction errors or local variations in the point clouds. As a result, the estimated rotation and translation parameters are more strongly influenced by the most geometrically stable road markings.

The resulting transformation matrix is saved as a text file and applied to the target point cloud using a Python script. Matrix operations are handled using the NumPy library (Harris et al. 2020), while Laspy library (G. V. Horn, Hruska, and contributors 2023) is used to load, manipulate, and write the point cloud data.

3.8 Alignment Accuracy Evaluation

To assess the quality of the alignment between the reference and target point clouds, this research applies a raster-based elevation comparison method known as DTM differencing, following the approach described in Salach et al. 2018. This method evaluates how well the aligned datasets match by comparing their ground surface elevations, represented as DTMs, under the assumption that there has been minimal ground surface change during the time gap between acquisitions.

Direct measurement of horizontal alignment errors is possible by comparing the positions of individual features. However, this requires the presence of other clearly defined and consistently extracted features in both datasets, besides the road markings, which may not always be available or reliable. Since road markings are already used as the control features for the alignment process, their use as independent check points would introduce bias into the evaluation. Therefore, DTM differencing is another way to measure the alignment accuracy without relying on the same features used for alignment.

Although this method mainly measures vertical differences, converting the point clouds into a regular grid helps reduce the impact of small horizontal shifts. If the horizontal error is smaller than the raster cell size, it has little effect on the elevation values because the points are averaged within each cell. This makes it easier to focus on overall alignment quality. DTM differencing also produces a large number of comparison points (raster pixels), which allows for more robust statistical analysis of the alignment results. For these reasons, it is a practical method for large-scale validation, especially when other feature-based checks are not available and the ground surface is expected to stay mostly the same.

The process begins by applying ground filtering to both the transformed and untransformed point cloud data, retaining only the points classified as ground. If the point cloud does not contain predefined ground classifications, ground filtering is performed using the CSF plugin (Zhang et al. 2016) in the CloudCompare software. This step ensures that the resulting DTMs represent the bare-earth surface, free from interference from buildings, vegetation, or other elevated features.

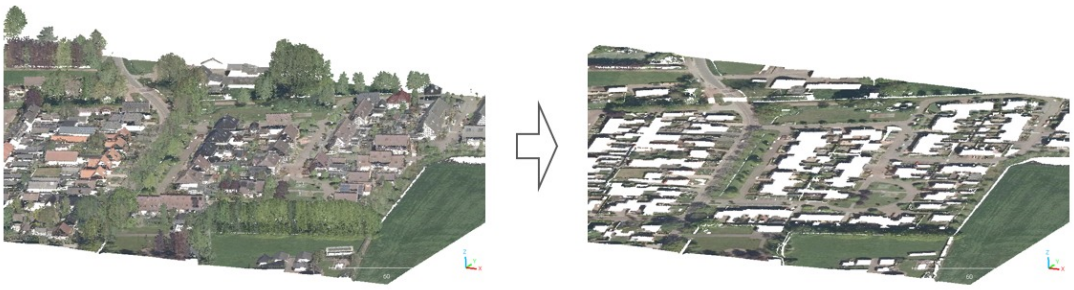


Figure 3.23: Illustration of ground filtering

Following ground filtering, each filtered point cloud is converted into a raster using a Python programme. The output raster file is written using the rasterio package (Gillies and contributors 2013), the input point cloud data are read using the laspy package (G. V. Horn, Hruska, and contributors 2023), and the raster values are computed using numpy (Harris et al. 2020). A regular raster grid is defined based on a user-specified resolution, and the elevation value for each cell is calculated as the mean of all ground point elevations falling within that cell. If no points fall within a given cell, a NoData value is assigned. This produces a continuous elevation model suitable for surface comparison.

To enable direct comparison between the reference and target DTMs, the target DTM is resampled to match the spatial resolution, extent, and grid alignment of the reference DTM. The resampling is carried out using a Python programme, employing bilinear interpolation through the reproject function provided by the rasterio package (Gillies and contributors 2013). Ensuring spatial consistency between the target and reference raster datasets allows the elevation values of the target DTM to be subtracted from those of the reference DTM on a per-pixel basis.

After point cloud alignment, the elevation differences are computed by subtracting the resampled target DTM from the reference DTM. The resulting raster represents vertical discrepancies across the study area. In well-aligned datasets, elevation differences should be close to zero, while larger positive or negative values may indicate misalignments or inconsistencies.

The decision to use only ground features when measuring differences between overlapping datasets is based on the assumption that the ground surface is the most stable element in the landscape. In the context of multi-temporal data, stable refers to features that remain largely unchanged over time. Unlike trees, which grow, or buildings, which may be constructed, altered, or demolished, the ground typically exhibits minimal variation. However, some low vegetation such as crops, grasses, bushes, or piles of leaves may be mistakenly classified as ground due to their low elevation, making them hard to distinguish from true ground points. In more extreme cases, groundworks such as excavation or landfilling for construction or agricultural activities can also alter the ground surface significantly. These factors can cause small changes in ground height that appear as high difference in the DTM difference results between different datasets.

To minimise the influence of the extreme ground change due to non-natural work and ensure a more reliable assessment of alignment accuracy, an interquartile range (IQR) filter is

applied to the DTM difference raster. The IQR filter addresses this by retaining only the middle value of the data, thereby excluding unusually high or low values. This filtering step helps focus the evaluation on meaningful discrepancies due to misalignment, rather than unrelated surface changes.

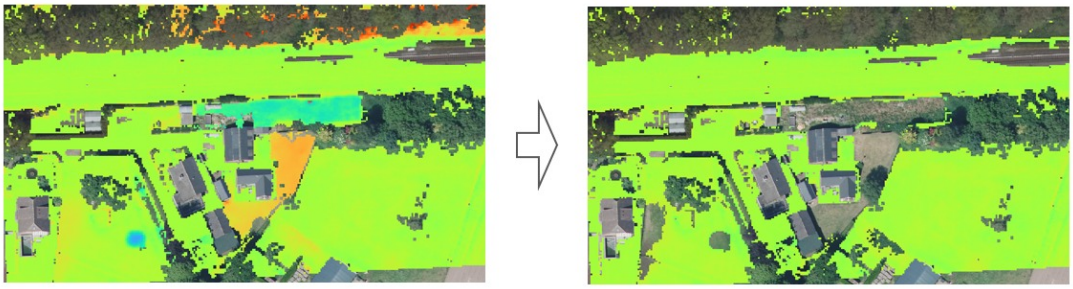


Figure 3.24: Illustration IQR filtering to DTM values. Some irrelevant pixel values due to change in environment are removed

Finally, the filtered DTM difference raster is exported, and key summary statistics such as the minimum, maximum, mean, and standard deviation are computed. These metrics provide a quantitative measure of alignment accuracy and are used to compare with different alignment results.

3.9 Alignment Results Comparison

Three result comparison schemes are used to explore the following: (1) the impact of reducing the influence of geometrically unstable road markings; (2) the effect of automatic road marking extraction quality on alignment accuracy; and (3) the variation in alignment performance as the distance increases from the area containing the road markings.

3.9.1 Approach to Evaluate the Influence of Geometric Stability Filtering on Alignment

The first comparison scheme aims to evaluate the impact of reducing the influence of geometrically unstable road markings by applying RANSAC inlier count-based weighting in the centroid alignment process. It is based on the assumption that counting on the unstable road markings, whether caused by extraction errors, changes in the condition or shape of the markings between the reference and target datasets, or differences in sensor specifications that cause the markings to be shifted or poorly represented, the same as the stable road marks can reduce alignment accuracy.

The comparison is carried out by evaluating the alignment results obtained using the RANSAC-Weighted Centroid Alignment method proposed in Section 3.7, against those obtained using a more straightforward approach, such as simple centroid alignment that does not consider road marking stability during the alignment process.

3.9.2 Approach to Evaluate the Impact of Automatic Extraction Quality on Alignment Accuracy

The second comparison scheme aims to examine how the quality of road marking extraction, which was previously evaluated in Section 3.5, affects alignment accuracy. Manually digitised road markings are also used as a benchmark to represent the best possible alignment result, assuming the road markings are perfectly extracted. The underlying assumption is that lower extraction quality and detection accuracy will lead to a greater difference between the alignment results obtained from manually digitised and automatically extracted road markings.

The comparison is carried out by subtracting the DTM difference of the alignment result using automatically extracted road markings from the DTM difference of the alignment result using manually digitised road markings, in order to obtain the DTM difference residuals. These residuals, calculated for each alignment case, are then compared against the extraction quality metrics to assess their relationship.

3.9.3 Approach to Evaluate Alignment Accuracy Decay with Increasing Distance from Road Markings

The third comparison scheme aims to examine how alignment errors originating from the benchmark area propagate to surrounding areas. It is based on the assumption that the further an area is from the road marking benchmark, the greater the potential impact of alignment errors. This may occur in scenarios where road markings are concentrated in a limited zone, such as urban centres or highway segments, while the surrounding areas lack sufficient road markings to serve as benchmarks. This scheme provides insights into the extent to which the current automatic road marking extraction method can be relied upon as a benchmark for accurately co-registering point cloud data across a larger area.

The comparison is carried out by dividing the DTM difference map into multiple segments based on increasing distance from the benchmark area. For each segment, the DTM difference values are extracted, and their average pixel values are computed to represent the local alignment error. These average values are then compared across segments to observe how alignment accuracy changes as the distance from the road marking benchmark increases. This approach allows for the assessment of how far the alignment transformation, derived from the benchmark area, remains reliable.

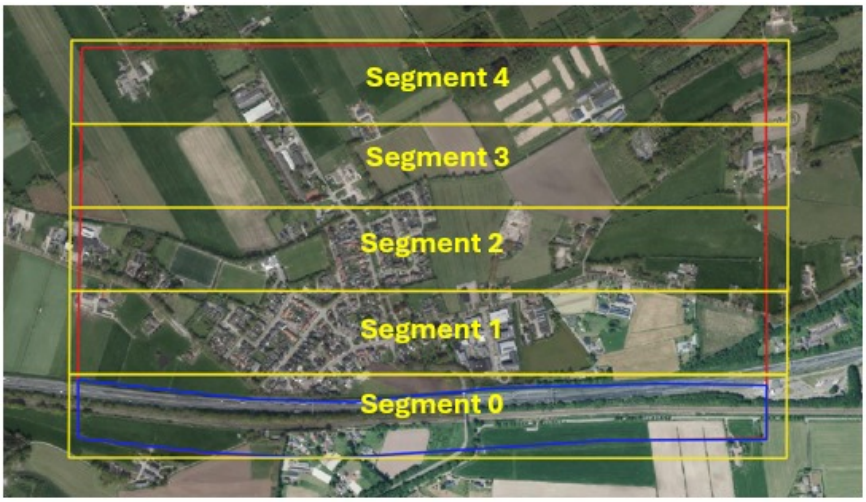


Figure 3.25: DTM difference map divided into multiple segments. The red polygon indicates the point cloud AOI, the blue polygon marks the area containing road markings used as the benchmark, and the yellow polygons represent the individual DTM segments.

Since the extended AOI is not fully covered by the Prorail dataset, only the AHN point clouds are suitable for this analysis. To investigate how alignment accuracy changes with increasing distance from the road marking benchmarks, alignments are performed between different AHN datasets using road markings as control features. In this setup, it is assumed that the road markings used for alignment are located only within the southernmost segment. The remaining segments used for evaluation are arranged to the north of this benchmark area, allowing a systematic assessment of how well the alignment holds as the distance from the control features increases in the northward direction.

4 Results

This chapter presents the results of the alignment comparisons, following the methodology outlined in Chapter 3, to address the research questions introduced in Chapter 1. The results provide the foundation of analysis that used to answer the main research question regarding the suitability of using road markings extracted using the method proposed in Heide 2024 as benchmarks for heterogeneous LiDAR point cloud harmonisation.

The results are organised according to the corresponding sub-research questions. Section 4.1 presents the results of the correspondence establishment and the subsequent alignment using the established correspondences. Section 4.2 discusses the impact of weighting road markings based on their geometric stability to validate the proposed method. In Section 4.3, the alignment results comparison between automatically extracted road markings and manually digitised road markings is shown. Section 4.4 examines the effect of varying distances from the benchmark area on alignment accuracy. Finally, Section 4.5 analyses the potential sources of uncertainty in the alignment process using automatically extracted road markings, as observed throughout the research pipeline.

4.1 Correspondence Establishment Results

This section presents the results of the correspondence establishment performed across various alignment time gaps, using the approach described in Section 3.6. The proposed method involves two main steps applied to road markings in overlapping datasets. The first step filters out road markings that are isolated from any corresponding marking in the overlapping dataset, either due to false detection or environmental changes. The second step assesses the geometric stability of the remaining pairs in the context of point cloud alignment.

4.1.1 Established Correspondences

The first step in the correspondence establishment process aims to identify and remove road marking segments that lack a meaningful spatial match in the overlapping dataset. These rejected segments may result from false positives in the extraction process or genuine changes in the road environment between acquisition dates. Table 4.1 and Table 4.2 summarise the number of manually digitised and automatically extracted road markings, respectively, that were rejected during this initial filtering stage across several alignment time gaps.

Table 4.1: Total number of rejected manually digitised road markings in the initial filter through several time gaps of alignment

Manually Digitized Road Markings				
Reference	Target	Total target RM	Rejected in initial filter	Total processed RM
AHN3 (2018)	Prorail2019	292	3	289
	Prorail2021	316	200	116
	Prorail2023	320	209	111
AHN4 (2020-2022)	Prorail2019	292	66	226
	Prorail2021	316	130	186
	Prorail2023	320	172	148
AHN5 (2023)	Prorail2019	292	228	64
	Prorail2021	316	136	180
	Prorail2023	320	117	203

Table 4.2: Total number of rejected automatically extracted road markings in the initial filter through several time gaps of alignment

Automatically Extracted Road Markings				
Reference	Target	Total target RM	Rejected	Total processed RM
AHN3 (2018)	Prorail2019	559	240	319
	Prorail2021	369	306	63
	Prorail2023	696	588	108
AHN5 (2023)	Prorail2019	559	470	89
	Prorail2021	369	275	94
	Prorail2023	696	503	193

It is important to note that automatically extracted road markings were not available for the AHN4 (2020–2022) dataset at the time of analysis. As a result, the comparison in Table 4.2 only includes AHN3 (2018) and AHN5 (2023) as reference datasets.

From the results in Tables 4.1 and 4.2, a clear trend can be observed in both the manually digitised and automatically extracted road markings: a larger time gap between datasets generally results in a lower number of road markings passing the initial filtering step. This suggests that the geometric consistency of road markings tends to decrease over time due to physical changes in the road environment, data quality differences, or extraction inconsistencies. Furthermore, the number of rejected road markings is consistently higher in the automatically extracted dataset compared to the manually digitised one. This highlights the

limitations of automatic extraction methods, which are more likely to include road markings that are not stable, particularly when aligning datasets with greater temporal differences.

Another observation from Tables 4.1 and 4.2 is the consistently higher number of total target road markings in the automatically extracted dataset compared to the manually digitised one. This indicates a trend of over-detection in the automatic extraction method, where more features, including noise or objects other than road markings, are initially selected. As a result, some of these over-detected features may pass the initial filtering stage and be included in the next process. This is shown in the alignment of Prorail 2019 to AHN3 and Prorail 2019 to AHN5, where the total number of processed road markings in the automatic dataset is higher than in the manual one.

Following the initial filtering step, the second part of the correspondence establishment process focuses on evaluating the geometric stability of the remaining road marking pairs. This is done by applying the RANSAC algorithm to identify consistent correspondences across multiple alignment iterations. In this study, 1000 RANSAC iterations were used to ensure a robust estimation of inliers. A higher number of iterations increases the likelihood of capturing more reliable correspondences by exploring more transformation possibilities. During each iteration, 3 correspondences were randomly sampled to estimate the transformation. This sample size was chosen because it is the minimum number required to compute a rigid transformation in 3D space using line-based features B. K. P. Horn 1987

Tables 4.3 and 4.4 present the distribution of geometrically stable correspondences for each alignment scenario, using manually digitised and automatically extracted road markings, respectively. These inliers represent road markings that are not only detected in both datasets but also maintain a stable spatial relationship throughout the alignment iterations.

Table 4.3: Number of geometrically stable correspondences (RANSAC inliers) from manually digitised road markings across different time gaps of alignment

Manually Digitised Road Markings						
Reference	Target	RANSAC inlier count				
		0-200	201-400	401-600	601-800	801-1000
AHN3 (2018)	Prorail2019	133	5	9	11	131
	Prorail2021	84	0	2	1	29
	Prorail2023	84	1	3	1	22
AHN4 (2020-2022)	Prorail2019	86	4	3	11	122
	Prorail2021	60	3	6	11	106
	Prorail2023	94	1	2	3	48
AHN5 (2023)	Prorail2019	49	0	0	1	14
	Prorail2021	51	9	8	13	99
	Prorail2023	34	3	1	2	163

Table 4.4: Number of geometrically stable correspondences (RANSAC inliers) from automatically extracted road markings across different time gaps of alignment

Automatically Extracted Road Markings						
Reference	Target	RANSAC inlier count				
		0-200	201-400	401-600	601-800	801-1000
AHN3 (2018)	Prorail2019	155	82	2	80	0
	Prorail2021	43	0	1	1	18
	Prorail2023	81	2	1	0	24
AHN5 (2023)	Prorail2019	62	0	2	2	23
	Prorail2021	67	1	1	1	24
	Prorail2023	116	2	1	2	72

Tables 4.3 and 4.4 show the number of geometrically stable correspondences identified during the alignment process, based on manually digitised and automatically extracted road markings, respectively. The manually digitised road markings are used as the reference condition, representing an ideal extraction scenario without detection or line accuracy errors. As such, the distribution of RANSAC inliers from this dataset serves as a baseline for evaluating the quality of the automatically extracted road markings.

In all alignment scenarios, the number of stable correspondences in the highest inlier bin (801–1000) is consistently lower for the automatically extracted road markings compared to the manual benchmark. This indicates that the quality of automatic road marking extraction directly influences the number of geometrically stable pairs that can support alignment. Errors in detection and line geometry reduce the likelihood that extracted features will maintain consistent spatial relationships across datasets. As a result, even if the initial number of extracted road markings is high, many of them may not be reliable enough to contribute effectively to the alignment process. This suggests that improving the extraction process can increase the number of geometrically stable pairs to give more correspondence in the alignment process.

4.1.2 Alignment Results Using Established Correspondences

After the correspondences are established, the next step is to use them to calculate the transformation matrix in order to align the target point cloud to the reference point cloud, following the proposed method described in Section 3.7. The alignment results are then validated using the DTM differencing as explained in Section 3.8. The results of the alignment validation are shown in Table 4.5.

Table 4.5: Validation of alignment results using established correspondences

Mean DTM Difference (m) of Alignment Using Established Correspondence				
Reference	Target	Untranformed	Manually Digitised	Automatic Extracted
AHN3 (2018)	Prorail2019	0.060	0.012	0.014
	Prorail2021	0.002	-0.035	-0.001
	Prorail2023	0.012	0.007	0.010
AHN4 (2020-2022)	Prorail2019	0.105	0.000	-
	Prorail2021	0.047	0.004	-
	Prorail2023	0.054	0.024	-
AHN5 (2023)	Prorail2019	0.096	-0.006	-0.043
	Prorail2021	0.037	-0.011	-0.020
	Prorail2023	0.051	0.020	0.004

The alignment results for AHN4 using automatically extracted road markings are not included in this table, as the extracted road markings are not available at the time of analysis. Therefore, the alignment tests, including AHN4 as a reference, cannot be conducted in this research

From Table 4.5, it can be seen that in most cases, the manually digitised road markings perform slightly better than those from the automatically extracted road markings. This suggests that there are factors related to the automatic extraction method that may influence the alignment results, which will be discussed further in Section 4.3. In general, both road marking extraction methods successfully reduce the alignment gap in the overlapping data. However, in the alignment of Prorail2021 to AHN3 using manually digitised road markings, the alignment error increased. This is suspected to be caused by a limitation of the proposed alignment method, which will be further investigated in Section 4.2.

To answer the first sub-research question: correspondences between overlapping LiDAR datasets with different acquisition times can be established by applying a method that weights correspondences based on their stability. The results show that the established correspondences can be used to reduce the misalignment between overlapping point clouds. However, in some cases, they failed to do so, indicating a potential limitation of the proposed method. The number of reliable correspondences decreases as the time gap increases, mainly due to changes in the road environment. However, the number of stable correspondences from automatically extracted road markings is lower than that from manually digitised ones, indicating that there are factors in the automatically extracted road markings that affect the number of stable correspondences. Nevertheless, as shown in Table 4.5, the automatically extracted road markings still manage to reduce the misalignment between overlapping datasets.

4.2 Reducing the Influence of Unstable Road Markings on Alignment Accuracy

In Section 3.6, a method for establishing correspondences was introduced based on counting how often a pair is identified as an inlier across multiple alignment iterations. Building on this, Section 3.7 proposed a weighted approach for calculating the transformation matrix, where the inlier count for each pair is used as a weight in the alignment process. In Section 4.1, the results of the implementation of the method are shown.

In the following section, the result of the proposed method is validated by comparing it with a more straightforward approach for both correspondence establishment and transformation calculation as described in Section 3.9.1. The comparison method uses a simple nearest neighbour search to establish correspondences, similar to the initial filtering step described in Section 3.6, followed by transformation matrix estimation using the least-squares method introduced in Arun, T. S. Huang, and Blostein 1987, as reviewed in Section 2.4.3.

By comparing the results of both methods using mean DTM differences, as reviewed in Section 3.8, between aligned and untransformed point clouds, the aim is to evaluate whether weighting benchmark based on its geometric stability provides any practical advantage in improving alignment accuracy over this simpler baseline. The result of the validation of the method using manually digitised road markings is shown in Table 4.6, while the one that uses automatic digitised road markings is shown in Table 4.7.

Table 4.6: Comparison of RANSAC-Weighted and Nearest Neighbour Search using manually digitised road markings

Mean DTM Difference (m) Using Manual Digitised RM				
Reference	Target	Untransformed	RANSAC-Weighted	NN Search
AHN3 (2018)	Prorail2019	0.060	0.012	0.011
	Prorail2021	0.002	-0.035	-0.004
	Prorail2023	0.012	0.007	0.021
AHN4 (2020-2022)	Prorail2019	0.105	0.000	0.001
	Prorail2021	0.047	0.004	-0.005
	Prorail2023	0.054	0.024	0.015
AHN5 (2023)	Prorail2019	0.096	-0.006	-0.017
	Prorail2021	0.037	-0.011	-0.012
	Prorail2023	0.051	0.020	0.021

The results show that the difference between the RANSAC-weighted method and the simple nearest neighbour search method is very small. In most cases, the mean DTM differences between the two methods are only a few millimetres apart. Sometimes the RANSAC-weighted method gives slightly better results, and sometimes the nearest neighbour search method is better. The validation shows that adding weights based on geometric stability does not clearly improve the alignment compared to using a more straightforward nearest neighbour

Table 4.7: Comparison of RANSAC-Weighted and Nearest Neighbour Search using automatically extracted road markings

Mean DTM Difference (m) Using Automatic Extracted RM				
Reference	Target	Untransformed	RANSAC-Weighted	NN Search
AHN3 (2018)	Prorail2019	0.060	0.014	0.011
	Prorail2021	0.002	-0.001	-0.010
	Prorail2023	0.012	0.010	0.014
AHN5 (2023)	Prorail2019	0.096	-0.043	-0.024
	Prorail2021	0.037	-0.020	-0.005
	Prorail2023	0.051	0.004	0.008

search. This result is the same whether the road markings were extracted manually or automatically, and whether the time gap between datasets was short or long. Overall, the weighted method works, but it doesn't make a big difference in the final alignment accuracy. It is also important to note that these differences are measured over relatively small areas of interest. In larger AOIs, the impact of alignment errors may become more noticeable, and further evaluation would be needed to confirm if similar trends hold at broader scales.

To support the comparison between the RANSAC-weighted method and the nearest neighbour search approach, the following figures show selected alignment results using both methods. Each example includes a DTM difference map and the corresponding road marking distribution to highlight where spatial differences occur and how the placement and density of features may influence the alignment. These visuals help clarify whether the small differences in mean DTM values are spread evenly or caused by local misalignments. Only a few representative cases are shown to avoid redundancy, including one case where RANSAC performs better, one where the nearest neighbour method performs better, and one where both give similar results.

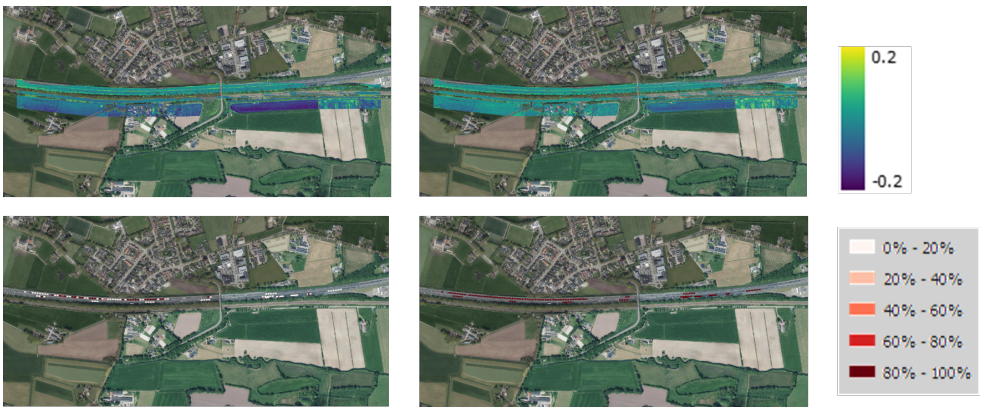


Figure 4.1: Case where Neighbour search method performs better than RANSAC-Weighted method. (a) RANSAC-Weighted. (b) Nearest Neighbour search.

Figure 4.1 shows the alignment results for the manually digitised road markings from Prorail2021 to AHN3 (2018), comparing the RANSAC-weighted method (left) and the nearest neighbour search method (right). The DTM difference maps indicate that the nearest neighbour approach produces slightly more uniform results (consistent green colour), while the RANSAC, weighted result shows more variation, particularly with bluish areas in the southern part of the DTM. The DTM difference statistic result is also shown in Table 4.6.

The lower performance of the RANSAC-weighted method in this case is suspected to be related to the distribution of road markings used. As shown in the bottom row, RANSAC selected fewer and more sparsely distributed road markings, with most concentrated on the northern lanes. Many of the markings on the southern lanes received low stability weights and were excluded (white-coloured). In contrast, the nearest neighbour method still retained some road markings in the southern lane. From this result, it is suspected that an uneven or linear distribution of benchmarks, such as road markings limited to one side of the road, can reduce alignment accuracy, especially in areas farther away from the available correspondences.

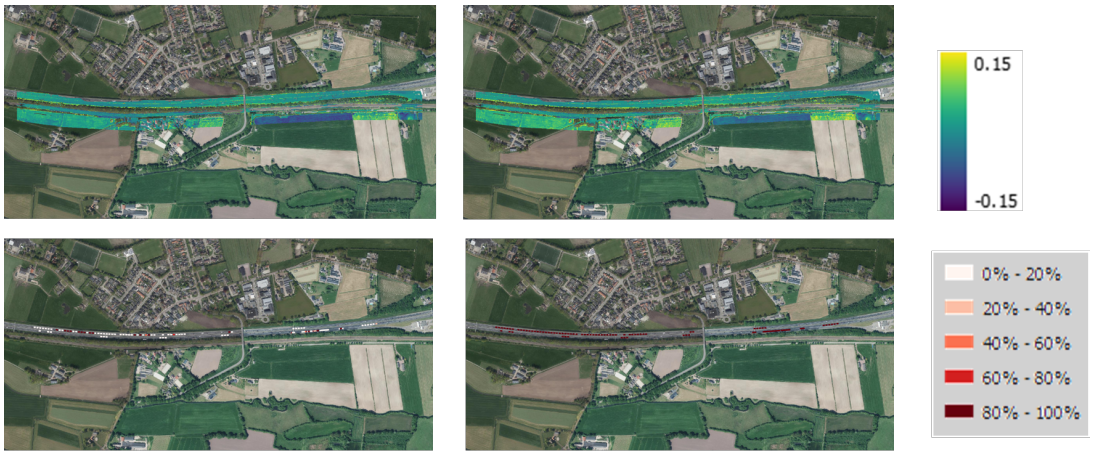


Figure 4.2: Case where RANSAC-Weighted method performs better than Nearest Neighbour search method. (a) RANSAC-Weighted. (b) Nearest Neighbour search.

Figure 4.2 shows the alignment results for the manually digitised road markings from Prorail2023 to AHN3 (2018), comparing the RANSAC-weighted method (left) and the nearest neighbour search method (right). The DTM difference maps indicate that the RANSAC-weighted method provides slightly more consistent results, which is shown by consistent green colour around the DTM. In contrast, the nearest neighbour method shows more variation, which can be seen by a slightly more yellow colour in the southern part of DTM. This is in line with the DTM statistics in Table 4.6.

The better performance of the RANSAC-weighted method in this case is surprising, considering it also uses fewer road markings with uneven distribution. As shown in the bottom row of Figure 4.2, most of the selected markings are located on the northern lanes, while many of the southern markings were excluded due to low stability weights. This observation suggests that the geometric stability of benchmarks plays a more important role than

their quantity in achieving accurate alignment. Although the RANSAC-weighted method discarded many southern road markings, the retained features offered a better alignment result. In contrast, the Nearest Neighbour method, by including all correspondences regardless of stability, may have introduced noise into the alignment process, especially in the northern lane where most of the available road markings are concentrated.

However, this also highlights a limitation of the approach used in this evaluation, as a further strategy is needed to quantify how the unstable road markings included in the Nearest Neighbour method reduce the quality of the alignment. At present, the analysis compares the overall performance of two different alignment methods, but it does not directly measure how much of the error is caused by the inclusion of low-stability features. This makes it difficult to confirm whether these features are the main reason for the reduced accuracy.

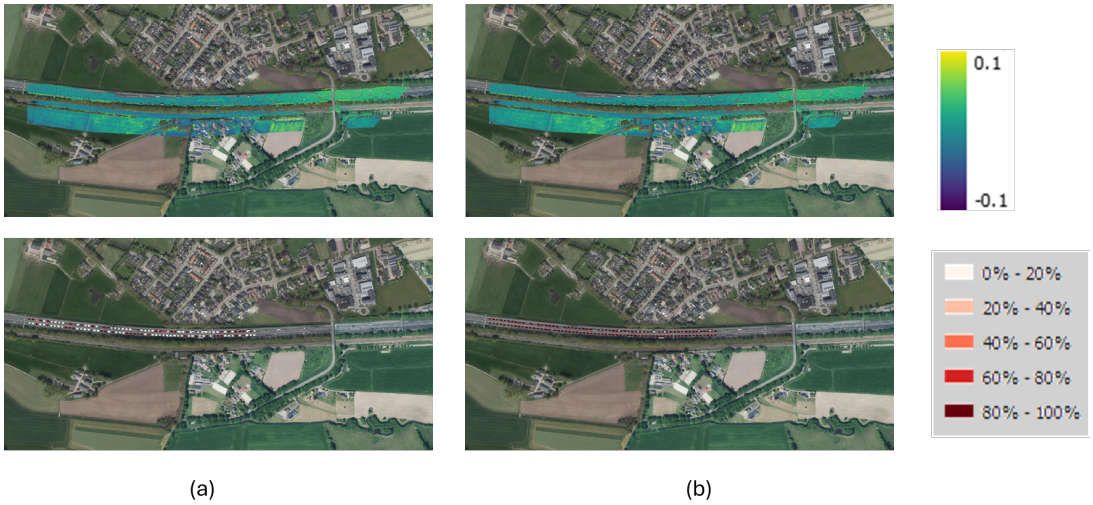


Figure 4.3: Case where RANSAC-Weighted and Nearest Neighbour search method perform equally. (a) RANSAC-Weighted. (b) Nearest Neighbour search.

Figure 4.3 shows the alignment results for automatically extracted road markings from Prorail2023 to AHN5 (2023), comparing the RANSAC-weighted method (left) and the nearest neighbour search method (right). The DTM difference maps show that both methods produce similarly good results, shown by consistent green colour for both DTM difference maps. This is supported by the values in Table 4.7, where the mean DTM difference is 0.004 m for the RANSAC-weighted method and 0.008 m for the nearest neighbour method.

In this case, the road marking distribution is relatively even across both lanes, as shown in the bottom row. This supports the idea that when road markings are well distributed across the benchmark area (in this case, a highway), both the RANSAC-weighted and nearest neighbour methods can perform equally well, and the choice of correspondence strategy becomes less critical.

To conclude the findings, the comparison of selected alignment cases illustrates that the performance of the RANSAC-weighted and nearest neighbour methods depends strongly on the distribution of the road marking correspondences. When road markings are unevenly distributed, especially limited to one side of the road, alignment accuracy may degrade,

especially in areas farther from the benchmark. For the RANSAC-weighted method, which relies on weighted inlier counts and tends to exclude less stable features, this issue is more likely to occur. However, as shown in the second case (see Figure 4.2), even with uneven distribution, the RANSAC-weighted method can still outperform the nearest neighbour approach. This outcome highlights a limitation of the evaluation approach in this research, although unstable road markings are likely to influence the alignment result, the method does not quantify their individual impact. Finally, when road markings are dense and evenly distributed across the alignment area, both methods produce almost equal results, indicating that the impact of the correspondence strategy becomes less significant. These findings highlight the importance of distribution when selecting or designing alignment benchmarks.

The influence of the time gap on the number and spatial distribution of usable road markings is influential in determining the robustness of the alignment. As shown in Section 4.1, larger time gaps lead to more road markings being rejected during the initial filtering step. This is mainly due to changes in the road environment over time, which cause many road markings to no longer be usable as benchmarks. This is also highlighted in Figure 4.2 and Figure 4.3, which both show the alignment of the Prorail2023 point cloud to different reference datasets (AHN3 2018 and AHN5 2023, respectively). From these figures, it can be seen that a smaller alignment time gap (Figure 4.3) results in a more complete set of candidate correspondences, making the RANSAC-weighted method less likely to over-filter the input features.

To answer the second sub-research question: the findings in this section show that selecting road markings based on geometric stability does not consistently lead to improved alignment. While the strategy can reduce alignment gaps and performs well when a sufficient number of well-distributed features are available, the actual improvement over a simpler nearest neighbour method is often minimal. In many cases, the difference between the two methods is only a few millimetres. This suggests that although the method helps to exclude correspondences with poor stability, it does not provide a significant advantage in alignment accuracy. The problem with this approach occurs when the number of initial correspondences is low or unevenly distributed, particularly in larger alignment time gaps, where changes to the road surface reduce the availability of reliable road markings. When the number of initial correspondences is low and unevenly distributed, the method tends to over-filter, reducing alignment quality by assigning low weight to spatially important road markings that help maintain an even distribution. Overall, the method is functional, but its added complexity does not always translate to better alignment results.

4.3 Impact of Extraction Quality on Alignment Accuracy

In Section 4.1, it was observed that there are differences in the number of initial correspondences between manually digitised and automatically extracted road markings. Furthermore, each extraction method results in a different composition of road marking stability. In addition, the two methods produce different alignment results. These observations indicate that there are factors within the extraction process that contribute to these differences. To better understand these factors, the quality of the automatically extracted road markings is assessed, then the alignment result is compared following the approach described in Section 3.9.2.

The following section investigates how the quality of automatically extracted road markings affects the final alignment accuracy in LiDAR co-registration. The alignment results from automatic extraction are compared with those from manually digitised road markings, which are assumed to give the best possible alignment.

4.3.1 Extraction Quality Measurement

This section presents the evaluation of the quality of automatically extracted road markings based on two aspects that were proposed in Section 3.5, which are: detection accuracy and extracted line accuracy. These measurements aim to assess how well the automatic method performs in identifying and representing road markings across different datasets.

Detection Accuracy

To evaluate how well the automatic extraction method identifies road markings, three standard detection metrics are used: completeness, correctness, and overall quality. These metrics are calculated by comparing the extracted road markings to manually digitised ground truth, across different datasets. The results are summarised in Table 4.8.

Table 4.8: Detection accuracy metrics for each dataset

Metric	Prorail 2019	Prorail 2021	Prorail 2023	AHN3	AHN5
True Positive	259	255	307	247	123
False Positive	300	115	390	352	100
False Negative	33	61	13	61	0
Completeness	0.887	0.807	0.959	0.802	1.000
Correctness	0.463	0.689	0.440	0.412	0.552
Quality	0.438	0.592	0.432	0.374	0.405

Overall, the results highlight a trade-off in the extraction method: while it successfully detects most actual road markings (high completeness), it also includes a significant number of incorrect detections (low correctness). Completeness is primarily influenced by false negatives, which occur when actual road markings are not detected despite being present. On the other hand, correctness is largely affected by false positives, where non-road features are mistakenly identified as road markings. These false positives arise from various issues, such as the inclusion of edge lines or noise points that are incorrectly modelled as road markings.

Completeness measures how many of the actual road markings in the data were successfully detected by the method. A high completeness score means that most of the real road markings were found, while a low score means many were missed. For example, in the Prorail 2021 and AHN3 datasets, the completeness scores are around 0.80, which means that about 20% of the road markings were not detected. This can happen for several reasons. Sometimes, the markings are missed because their intensity values fall outside the detection

range. In other cases, they may be removed due to filtering steps of the automatic extraction process if they do not meet the condition.

As shown in Table 4.8, the method generally achieves high completeness across all datasets, with the highest result found in AHN5 (1.0) and Prorail 2023 (0.959). Referring to Table 3.1, despite AHN 3 having a much lower density of just 10 points/m², it still achieves a completeness score of 0.802, which is comparable to Prorail 2021 (0.807), which has a density of 134 points/m². A similar pattern can be seen between Prorail 2023 and AHN5, where the former has a much higher density than the latter, yet both achieve similarly high completeness scores. On the other hand, the difference in completeness between Prorail 2021 and Prorail 2023, as well as between AHN3 and AHN5, despite relatively small changes in density, reaches nearly 0.2. No consistent trend between density and completeness can be identified from these comparisons. This suggests that point density alone does not directly determine detection performance.

In terms of intensity range, the results also do not show a clear or consistent relationship with completeness. Prorail 2021 has one of the widest intensity ranges (25,000 to 40,000) and achieves a completeness score of 0.807. In comparison, AHN3 has the narrowest intensity range (110 to 380) but still achieves a completeness score of 0.802, which is almost identical to that of Prorail 2021. Similarly, Prorail 2023 and AHN5 achieve almost similar completeness scores of 0.959 and 1.000, respectively, even though their intensity ranges differ significantly; Prorail 2023 has a broader range (27,000 to 36,000), while AHN5 has a much narrower range (1,230 to 4,400). These results suggest that the intensity range alone does not explain the variation in completeness across datasets. However, despite the variation in intensity ranges, the automatic extraction method manages to achieve detection accuracy between 0.8 and 1.0, indicating that nearly all road markings recorded in the point cloud were successfully extracted.

Apart from completeness, another important aspect of detection performance is correctness. Correctness measures how many of the detected road markings are actually correct. For example, in the Prorail 2023 dataset, the correctness score is only 0.44, and for AHN3 it is 0.41. This means that more than half of the detected features are actually not road markings. This problem happens because the method often selects extra points by mistake, such as parts of nearby objects or noisy points that look similar in intensity. Even though the method is good at finding most of the real road markings (high completeness), it also includes a lot of wrong ones. This suggests improvement on the filtering step to avoid picking up features that do not belong to road markings.

Based on the results shown in Table 4.8, no clear or consistent relationship can be observed when comparing correctness with the intensity range and point density values listed in Table 3.1. For example, AHN3 has the lowest point density (10 points/m²) and a very narrow intensity range (110 to 380), yet its correctness score (0.412) is similar to that of Prorail 2023 (0.440), which has a much higher point density (170 points/m²) and a broader intensity range (27,000 to 36,000). Likewise, Prorail 2021 achieves the highest correctness score (0.689), despite having intensity and density values that are not significantly different from the other Prorail datasets. These comparisons suggest that correctness is not directly influenced by point density or intensity range alone.

Source of Error in Completeness Metric

To better understand the cause of these missed detections, a closer inspection was conducted on a manually prepared sample of the Prorail 2021 dataset, as illustrated in Figure 4.4. This visual comparison highlights areas where visible road markings were not captured by the automatic extraction. Building on this observation, the following discussion considers several possible explanations based on the steps in the automatic extraction pipeline (see Table 2.3).

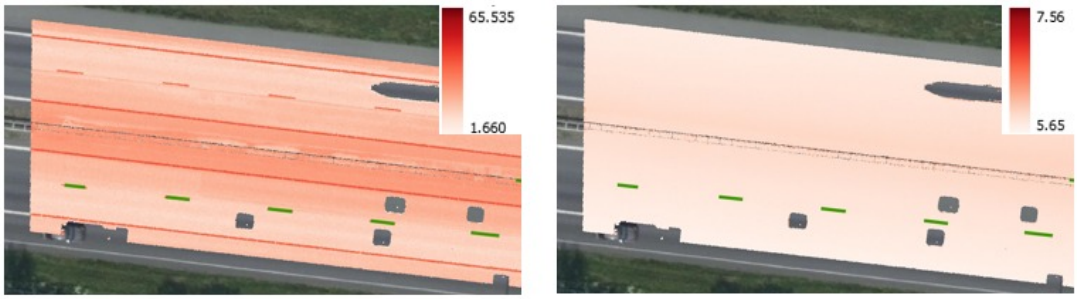
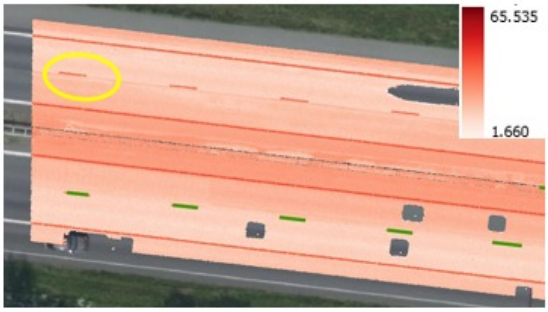


Figure 4.4: Example of undetected road markings in Prorail 2021 Data. Green lines represent extracted road markings. Viewed through (a) intensity value and (b) elevation value

The point cloud shown in Figure 4.4 has been manually ground classified and cropped to retain only the road surface, which follows key preprocessing steps used in the automatic road marking extraction pipeline of Heide 2024. This was done to ensure a fair comparison between the manually visualised data and the automated process. By applying similar filtering steps, it can be assumed that the displayed intensity values are similar to those used during automatic extraction, allowing for a more accurate investigation of why certain road markings were not detected.

Referring to the automatic road marking extraction pipeline proposed by Heide 2024 reviewed in Section 2.2, several possible reasons were considered to explain the undetected road markings in Prorail 2021 data shown in Figure 4.4. The first possibility is that the adaptive intensity thresholding step was too strict. However, when the point cloud is viewed through the intensity values (Figure 4.4a), the undetected road markings appear to have similar intensity values as the ones that were successfully extracted. If the thresholding were truly too strict, other markings would have also been excluded. The next possibility is that the markings may have been occluded by obstacles such as vehicles. Yet, this is also unlikely, since the undetected road marking points are detected in the intensity view.

Another consideration is that the points were mistakenly removed as part of the curb and noise segmentation step, under the assumption that they belonged to curb structures. However, as shown in the elevation view (Figure 4.4b), the elevation in this area is flat and consistent, indicating no significant height changes that would suggest curbs. To determine whether the road markings were undetected because they were rejected by the minimum and maximum geometric thresholds, an undetected road marking was segmented, and its geometric features were calculated. These values were then compared with the defined minimum and maximum threshold ranges, as illustrated in Figure 4.5.



(a)

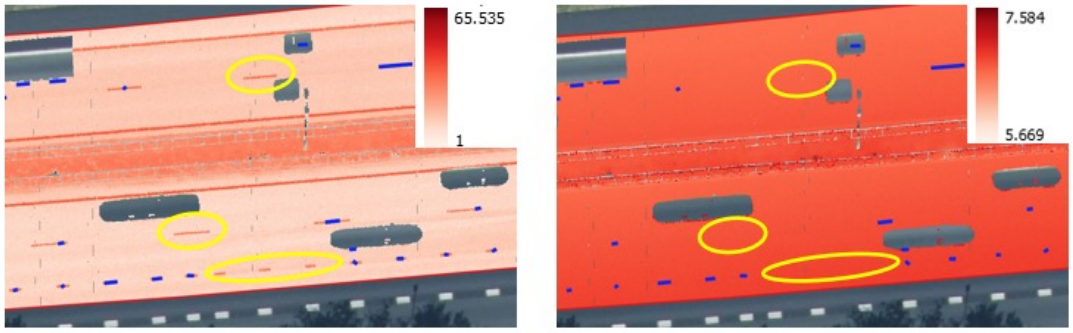
Geometric Features	Value
Linearity	0.9862
Planarity	0.0137
Sphericity	0.0001
Anisotropy	0.9999
Sum of Eigenvalues	0.7454
Change in Curvature	0.0001

(b)

Figure 4.5: Geometry features of an undetected road marking. (a) Location of road marking where the geometry values are calculated, symbolised by the yellow circle. (b) Calculated geometry features of selected undetected road markings.

By comparing Figure 4.5b with Table 3.2, it can be seen that the geometric feature values of the selected undetected road marking fall within the minimum and maximum threshold ranges. This analysis indicates that the road marking segment was not removed due to being outside the defined geometric thresholds.

Based on the reviews, none of the suspected causes, such as strict intensity thresholding, occlusion by obstacles, misclassification as curbs, or rejection by geometric thresholds, can fully explain why the road markings were missed. Each of these factors was investigated and ruled out through visual and numerical inspection. This suggests that the cause of the missed detection may lie in the later stages of the extraction pipeline that have not yet been reviewed in detail, specifically Step 5 (line segmentation) and Step 6 (zigzag and direction correction) of the automatic road marking extraction method used in this research, according to Table 2.3.



(a) (b)

Figure 4.6: Example of undetected road markings in Prorail 2023 Data. Viewed through (a) intensity value and (b) elevation value

The problem of undetected road markings also appears in the Prorail 2023 dataset, as shown in Figure 4.6. Even though this dataset achieved the highest completeness score (0.959) in

Table 4.8, a few markings were still missed, possibly for the same reasons as in Prorail 2021. The point density (155 points/m²) and intensity range (21 to 35,000) in Prorail 2023 are close to those in Prorail 2021, which had a lower completeness score. However, in Prorail 2023, this issue caused only a small number of missed markings, while in Prorail 2021, it had a bigger effect. This shows that the error does not follow a clear pattern and may also be influenced by other factors that were not covered in this analysis.

To conclude, while the automatic road marking extraction consistently achieves high completeness across datasets, the score is still not perfect due to several reasons. As detailed through the visual inspection, some road markings were missed even though they appeared visually clear and met the expected intensity and geometric conditions. Several potential causes, such as overly strict intensity thresholding, occlusion by vehicles, misclassification as curbs, or rejection based on geometric feature thresholds, were investigated and ruled out through visual inspection. This indicates that the missed detections may originate from later stages in the automatic extraction pipeline that were not fully examined with this visual inspection approach, particularly Step 5 (line segmentation) and Step 6 (zigzag and direction correction), as listed in Table 2.3. These steps might introduce errors by removing valid road marking segments, which reduce the completeness score.

Source of Error in Correctness Metric

Similar to the previous step, a visual interpretation approach was used to inspect samples of false positives to understand several reasons underlying the errors. This involved overlaying the automatically extracted road markings and the manually digitised reference road markings on top of a preprocessed point cloud. The preprocessing step included cropping the point cloud to retain only the road surface and applying ground classification to remove non-ground objects. Additionally, the intensity range was adjusted to make the road markings visually stand out, helping to identify where detection errors had occurred.

The inspection was conducted using the AHN3 dataset, as it recorded the lowest correctness score among all datasets. Starting with this dataset helps focus on the worst case, where the problem of false detection is most obvious and easier to learn from.

Figure 4.7 shows a snippet of the AHN3 dataset overlaid with the manually and automatically extracted road markings; it can be seen that many road markings have been extracted using the automatic method. However, a significant portion of these do not correspond to the manually digitised road markings, as illustrated by the blue lines. Only the markings that overlap with the manually digitised ones, shown in red, represent correct detections. This shows that some of the detected road markings are incorrect. To better understand the types of false detection, a closer visual inspection of selected segments is carried out.

As shown in Figure 4.8, edge lines along the sides of the road, which are not intended to be included as part of the road markings used for co-registration benchmarks, are also extracted by the automatic method. These detections are counted as false positives, which lowers the overall correctness score. Moreover, the detection of the edge lines is not continuous; instead of being identified as a single long line, they appear as a series of short, dotted segments. This is likely due to the natural variation in the intensity values along the edge lines. As a result, the number of false positives increases further, leading to an additional decrease in the correctness score.

Figure 4.7: Detection error in AHN3 point cloud. Blue lines represent automatically extracted road marks, red lines represent manually digitised road marks (as a reference for detection accuracy measurement)

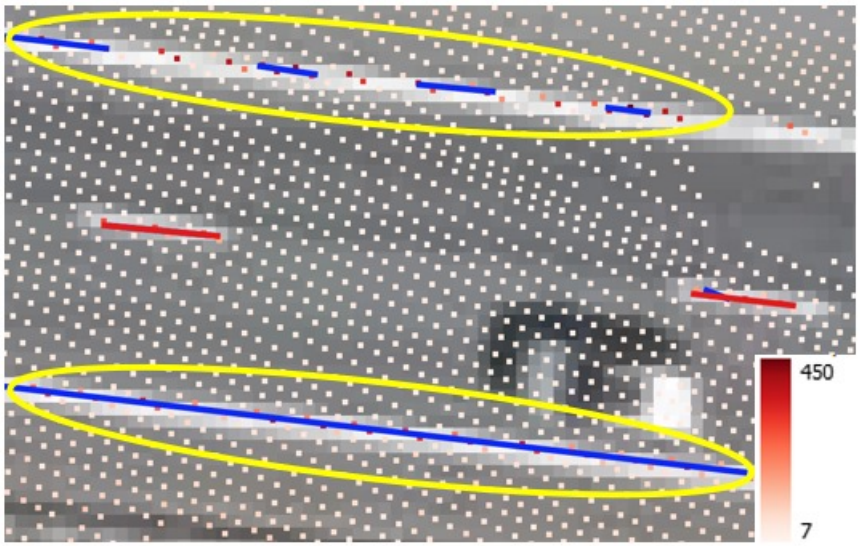


Figure 4.8: Closer inspection of false detections on edge lines. Blue lines are features extracted from the automatic extraction method, red lines are manually digitised road markings, and yellow-circled lines indicate the sample of false detection.

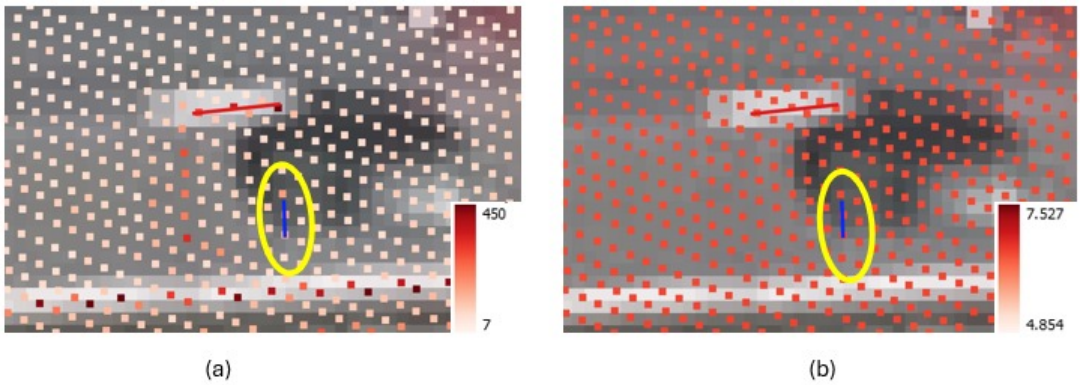


Figure 4.9: Closer inspection of false detections on noise points. The yellow-circled line shows a sample of false detection. (a) Viewed using intensity values, which show significantly high intensity. (b) Viewed using elevation values, which show no significant difference compared with the surrounding ground points.

Another type of false detection is caused by incorrect detection of undefined ground points with high intensity values, as shown in Figure 4.9. This is suspected to originate from points that were part of a car object (which typically produces high intensity values) but happen to have an elevation very close to the ground. As a result, these points may not have been properly filtered out during the vehicle removal step. When the distribution pattern of these points resembles that of a road marking, the automatic extraction algorithm may incorrectly classify them as a road marking segment and model them as a line. This leads to false positives and further reduces the correctness score.

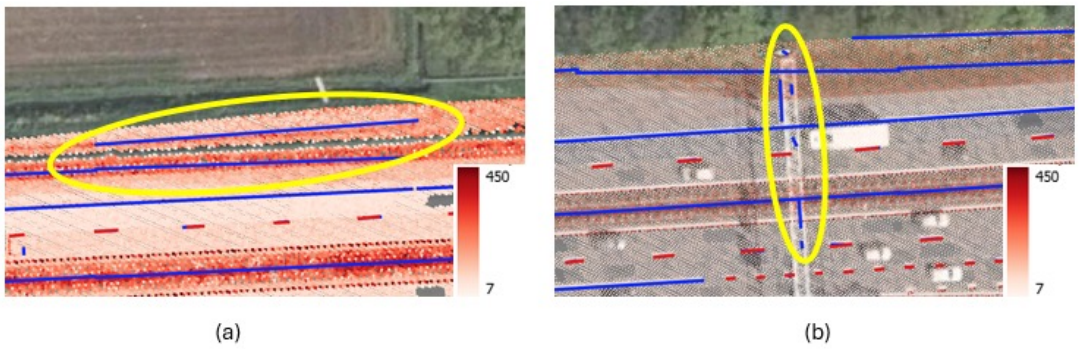


Figure 4.10: Closer inspection of false detections without clearly visible high-intensity point clusters. The yellow-circled areas show automatically extracted lines where no obvious group of high-intensity points is visible.

Another type of false detection happens when the extraction algorithm creates a road marking line in an area where there is no clear group of high-intensity points, as shown in Figure 4.10. In the yellow-circled area, there are no obvious signs of a real road marking, but the algorithm still detects and draws a line. This type of false detection may reflect a limitation of the inspection method itself. While no distinct segment is apparent through

visual analysis, it is possible that the structured steps of the automatic extraction pipeline identified a valid pattern that is not easily noticeable by eye. Alternatively, the error might be caused by leftover non-ground or noisy points with high intensity that were not fully removed during earlier filtering steps of the automatic extraction pipeline. These possibilities show that not all false detections can be easily explained just by looking at the data. Some of them might happen because of small steps in the automatic process that are hard to see or understand through simple visual inspection.

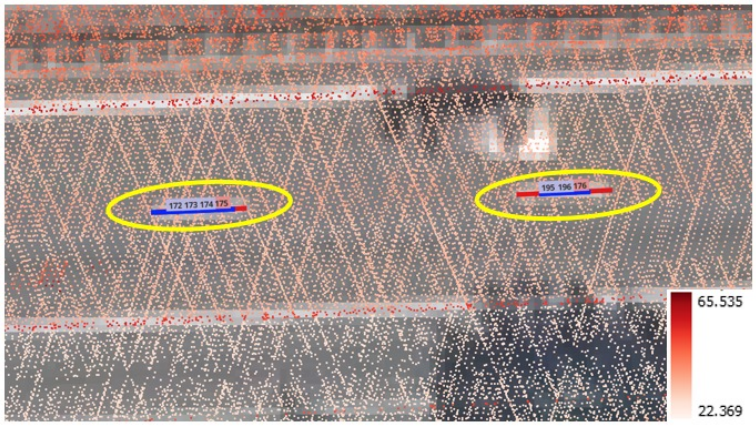


Figure 4.11: Closer inspection of false detections showing duplicated road marking extraction. The yellow-circled areas show cases where a single road marking is mistakenly extracted as two separate lines (blue lines), despite there being no clear indication of multiple point clusters in the intensity view.

A different type of false detection is observed in the Prorail datasets, where a single road marking is mistakenly detected multiple times. As shown in the figure, there is only one actual road marking present, yet the algorithm generates two separate line segments at the same location. However, the reason for this is not clearly explained through visual inspection, as the intensity view shows a continuous distribution of points that appear to form a single segment. When this happens, the extraction method treats these groups as separate features and generates multiple lines for what should be a single road marking. These duplicated detections are counted as false positives, which reduces the overall correctness score.

To conclude, the visual inspections revealed several factors that cause false positives. First, edge lines along the road, which are not intended as road markings for alignment, are also extracted due to their high intensity, even when their appearance is inconsistent. Second, noise points such as parts of vehicles or undefined objects, which have high intensity values and are mistakenly identified as road markings, especially if not fully removed during the filtering process. Third, some detections occur even where there is no clear intensity cluster, suggesting the algorithm might be responding to unidentified patterns in the data. These three cases show that false positives are not caused by a single factor, but by a combination of intensity, geometry, and insufficient filtering. Improving the filtering stages or adjusting the criteria used in the automatic method may help reduce these false detections and improve the correctness score.

Extracted Line Accuracy

To evaluate how well the lines were extracted using the automatic road marking extraction method, this section compares the geometry of the automatically extracted line features to manually digitised ground truth lines. The comparison looks at differences in position, length, and orientation. The results of the extracted line accuracy measurement are shown in Table 4.9.

Table 4.9: Extracted line accuracy metrics across datasets

Metric	Prorail 2019	Prorail 2021	Prorail 2023	AHN3	AHN5
Mean Δ Length 3D	-1.373	-1.316	-1.183	-0.984	-1.197
Mean Δ Angle Horizontal	4.542	5.490	4.860	6.858	4.995
Mean Δ Angle Vertical	4.497	4.760	3.475	1.300	1.444
Mean ΔX of Centroid	-0.028	-0.006	0.001	-0.020	0.008
Mean ΔY of Centroid	0.007	-0.004	-0.003	0.008	0.019
Mean ΔZ of Centroid	-0.001	0.005	0.000	0.001	0.000
Mean Δ3D of Centroid	**0.138**	**0.129**	**0.151**	**0.207**	**0.202**

Since the alignment method proposed in this research is the RANSAC-weighted centroid alignment method, which utilises the centroid of the extracted road markings to measure the transformation matrix needed to align the target point cloud to the reference point cloud (see Section 3.7), the main focus of the analysis is the mean 3D centroid difference. This metric shows how far the centroids of the extracted lines are from the centroids of the reference lines in 3D space.

The variation of density across datasets in Table 3.1 shows a relationship with the Mean 3D Centroid accuracy. A trend can be observed where datasets with higher point density tend to achieve better centroid accuracy. For instance, the Prorail datasets, which have point densities of 70, 144, and 170 points/m² respectively, show lower centroid differences (0.138, 0.129, and 0.151 metres). In contrast, the AHN3 and AHN5 datasets, with lower point densities of 10 and 17 points/m², show higher centroid differences of 0.207 and 0.202 metres.

Source of Error in 3D Centroid Position

To better understand the sources of error that lead to slight differences in the centroid positions between automatically extracted road markings and manually digitised road markings, a closer inspection was carried out on selected samples. Similar to the approach used in Section 4.3.1, this inspection was conducted visually by overlaying the automatically extracted and manually digitised road markings on a point cloud that had been ground-filtered and cropped to retain only the road surface.

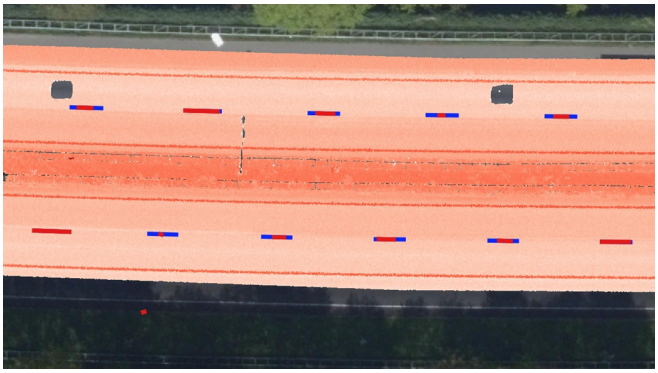


Figure 4.12: Sample of road markings detected in Prorail 2023 point cloud. Blue lines represent manually digitised road markings. Red lines represent automatically extracted road markings.

The sample of automatically extracted road markings and manually digitised road markings on the point cloud from the Prorail 2023 dataset is shown in Figure 4.12. In the figure, the blue lines are the manually digitised road markings (used as reference), and the red lines are the results from the automatic extraction. In an ideal case, the red lines should cover the blue lines completely, showing that the extracted road markings are in the correct position. In the figure, some red lines do match the blue lines well, but others only partly cover them or are slightly shifted. This indicates that there are variations in the accuracy of the extracted lines, which contribute to the overall mean error reported in Table 4.9. A closer investigation of several road marking samples with varying levels of extraction accuracy was conducted to understand the factors that cause these differences in how well the extracted road markings match the reference.

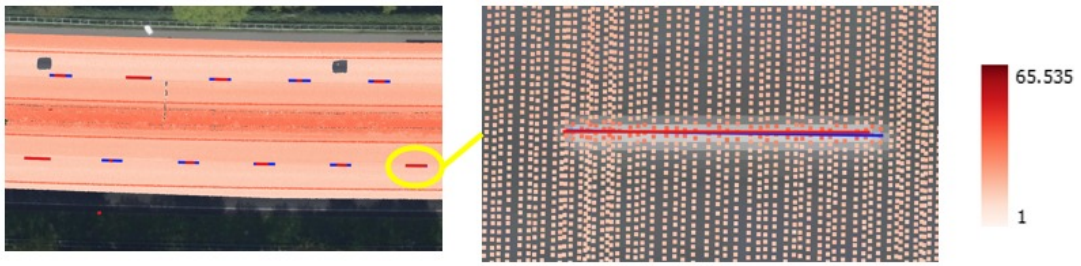


Figure 4.13: Example of a correctly detected road marking. Red lines represent automatically extracted road markings. Blue lines represent manually digitised road markings.

From Figure 4.18, it can be observed that the automatically extracted road marking (red line) closely matches the manually digitised reference (blue line) in both position and orientation. The red and blue lines are nearly overlapping, indicating that the extraction method successfully captured the correct location and shape of the road marking. The intensity values of the surrounding points also appear consistent, further supporting the correctness of the detection. This example illustrates a case of high geometric accuracy, where the centroid difference between the extracted and reference line is minimal. Cases like this contribute positively to the overall centroid accuracy scores.

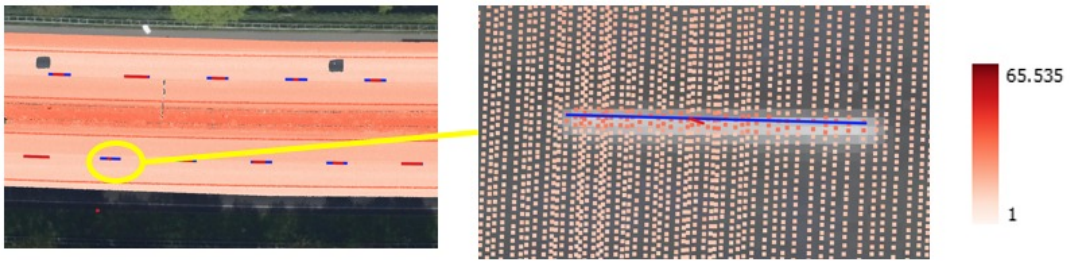


Figure 4.14: Example of a poorly detected road marking. Red lines represent automatically extracted road markings. Blue lines represent manually digitised road markings.

From Figure 4.14 it can be seen that the automatically extracted road marking (red line) differs from the manually digitised reference (blue line) in both length and orientation. The red line is noticeably shorter and slightly rotated compared to the blue line, indicating that the automatic extraction method did not fully capture the true shape and extent of the road marking. This kind of geometric mismatch contributes to centroid and angular errors.

Furthermore, this poor line extraction is not supported by the intensity values of the surrounding points. The visualised intensity clearly shows a continuous high-intensity segment corresponding to the road marking, which should have been detected as a complete and correctly oriented line. The fact that the algorithm failed to model it accurately. This suggests a potential error in the extraction pipeline, possibly in the clustering, line fitting step or line filtering.

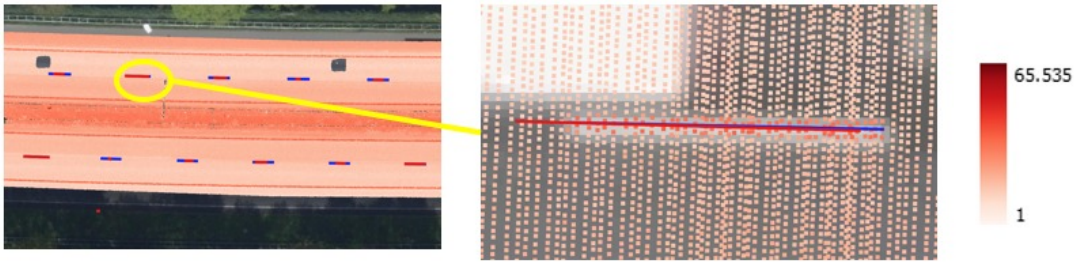


Figure 4.15: Example of a shifted detected road marking. Red lines represent automatically extracted road markings. Blue lines represent manually digitised road markings.

From Figure 4.15 it can be seen that the automatically extracted road marking (red line) has a similar length and orientation to the manually digitised reference (blue line), suggesting that the shape of the road marking was generally captured correctly. However, the extracted line appears slightly shifted to the right compared to the reference. This offset causes a difference in centroid position, which contributes to the centroid error.

Despite the surrounding intensity values clearly showing where the road marking is located, the extracted line does not align precisely with the centre of the visible high-intensity points. This suggests that there may be a limitation in the automatic extraction pipeline that makes the extracted line shifted from the original position, which cannot be detected or analysed through the proposed approach in this research.

To conclude, the inspection revealed two main types of errors that contribute to centroid inaccuracies. First, some road markings were extracted with incorrect shape or orientation, where the detected line was shorter or slightly rotated compared to the reference. This error may result from limitations in the clustering or line fitting steps that fail to fully capture the road marking's extent. Second, other road markings showed a noticeable shift, where the extracted line was offset from the high-intensity centre despite having the correct shape. The source of this error cannot be understood through the approach proposed in this research

4.3.2 DTM Difference Residuals Measurement

In Section 4.3.1, it was shown that the automatic extraction method introduces errors when compared to the manually digitised road markings. These errors come from both detection mistakes and inaccuracies in the shapes and positions of the extracted lines. In Section 4.1, it was shown that the quality of the extracted road markings directly affects how many stable pairs can be used to support the alignment process. Based on those earlier findings, this section now compares the alignment results using both types of road markings, manual and automatic, to see how the errors in the automatic extraction reduce the quality of the alignment compared to the manual benchmark.

To explore this further, Table 4.10, Table 4.11, and Table 4.12 present a summary of the key metrics used to evaluate both the extraction quality and the resulting alignment performance for each test. Each table includes measures of line centroid accuracy, detection completeness and correctness, the distribution of RANSAC inlier weights, and the resulting mean DTM residuals. These combined indicators provide a basis for analysing how different aspects of the automatic extraction process, such as geometric accuracy and the number of usable road markings, contribute to the overall co-registration result when compared to the manual benchmark

Table 4.10: Summary of extraction and alignment quality metrics for Test 1 and Test 2

		Test 1		Test 2	
		Reference	Target	Reference	Target
		AHN3	Prorail2019	AHN5	Prorail2019
Mean $\Delta 3D$ of Centroid (m)		0.207	0.138	0.202	0.138
Completeness (%)		0.802	0.887	1	0.887
Correctness (%)		0.412	0.463	0.552	0.463
Quality (%)		0.374	0.438	0.405	0.438
Weight	0% - 20%	155		62	
	20% - 40%	82		0	
	40% - 60%	2		2	
	60% - 80%	80		2	
	80% - 100%	0		23	
Mean Residual Map (m)		0.0022		-0.0485	

Table 4.11: Summary of extraction and alignment quality metrics for Test 3 and Test 4

		Test 3		Test 4	
		Reference	Target	Reference	Target
		AHN3	Prorail2021	AHN5	Prorail2021
Mean $\Delta 3D$ of Centroid (m)		0.207	0.129	0.202	0.129
Completeness (%)		0.802	0.807	1	0.807
Correctness (%)		0.412	0.689	0.552	0.689
Quality (%)		0.374	0.592	0.405	0.592
Weight	0% - 20%	43		67	
	20% - 40%	0		1	
	40% - 60%	1		1	
	60% - 80%	1		1	
	80% - 100%	18		24	
Mean Residual Map (m)		0.0395		-0.0116	

Table 4.12: Summary of extraction and alignment quality metrics for Test 5 and Test 6

		Test 5		Test 6	
		Reference	Target	Reference	Target
		AHN3	Prorail2023	AHN5	Prorail2023
Mean $\Delta 3D$ of Centroid (m)		0.207	0.151	0.202	0.151
Completeness (%)		0.802	0.959	1	0.959
Correctness (%)		0.412	0.44	0.552	0.44
Quality (%)		0.374	0.432	0.405	0.432
Weight	0% - 20%	81		116	
	20% - 40%	2		2	
	40% - 60%	1		1	
	60% - 80%	0		2	
	80% - 100%	24		72	
Mean Residual Map (m)		0.0031		-0.0156	

The results presented in Table 4.10, Table 4.11, and Table 4.12 show that there is no clear or consistent trend between individual extraction quality metrics (such as centroid accuracy, detection completeness, or the distribution of RANSAC inlier weights) and the resulting alignment accuracy, as measured by the mean DTM residuals. In some tests, higher completeness or more stable correspondences appear to correspond with better alignment results, but in

other cases, similar metric values produce different outcomes. This indicates that there is no direct correlation between any single quality measure and the final alignment accuracy. Therefore, evaluating extraction quality solely based on global statistics may not be sufficient to predict alignment success, and it highlights the need to include benchmark distribution inspection to better understand how extraction quality affects alignment accuracy, as it is also implemented in Section 4.2.

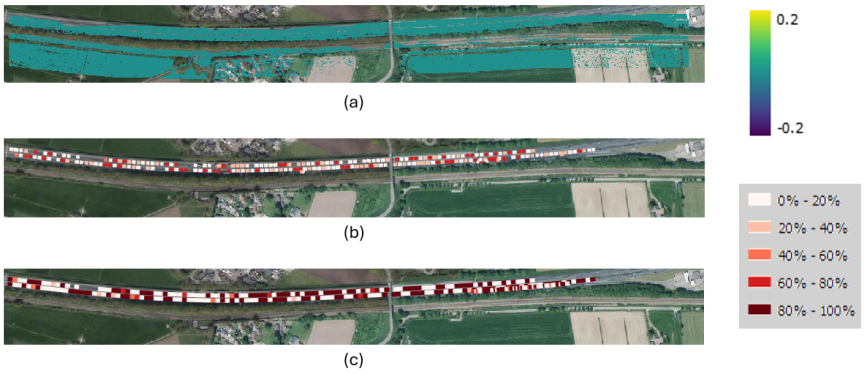


Figure 4.16: Test 1 benchmark distribution. (a) DTM residual map. (b) Automatically extracted benchmarks. (c) Manually digitised benchmarks

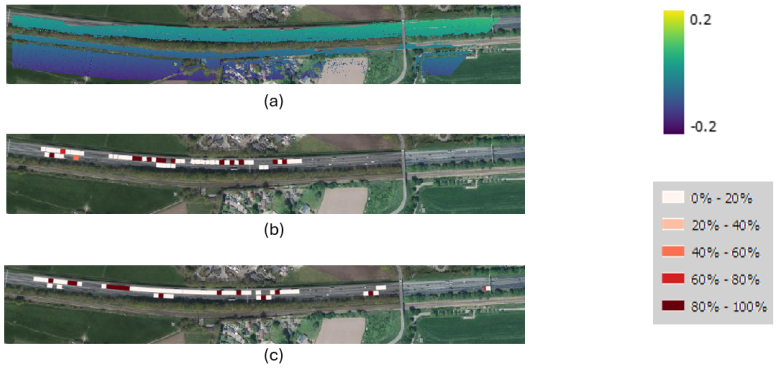


Figure 4.17: Test 2 benchmark distribution. (a) DTM residual map. (b) Automatically extracted benchmarks. (c) Manually digitised benchmarks

Based on the visual inspection of Figure 4.16 and Figure 4.17, the DTM residual map in Test 1 (Figure 4.16a) shows a more uniform result with mostly greenish tones, indicating consistent alignment. In contrast, the DTM residual in Test 2 (Figure 4.17a) displays a noticeable gradient, with greenish areas in the northern lane and bluish areas in the southern lane. This difference is likely related to the spatial distribution of road markings. In Test 1, both the automatically extracted (Figure 4.16b) and manually digitised (Figure4.16c) road markings are well distributed across both lanes. In Test 2, however, the number of stable road markings is visibly lower, especially in the southern lane. The automatically extracted benchmarks (Figure 4.17b) show no road markings on the southern side, while the manually digitised ones (Figure 4.17c) still contain a few remaining features. This lower and uneven distribution of

road markings in Test 2 shows a match with the less consistent alignment result shown in the DTM residual map. This supports the findings in Section 4.2, which indicate that larger alignment time gaps increase the likelihood of discrepancies in alignment accuracy between automatically extracted and manually digitised road markings.

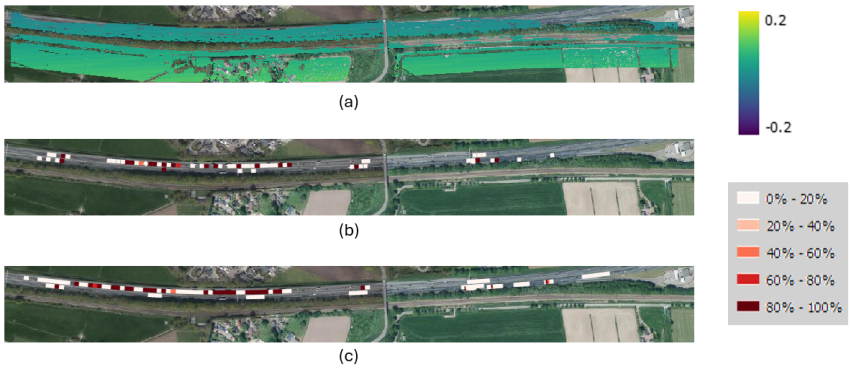


Figure 4.18: Test 3 benchmark distribution. (a) DTM residual map. (b) Automatically extracted benchmarks. (c) Manually digitised benchmarks

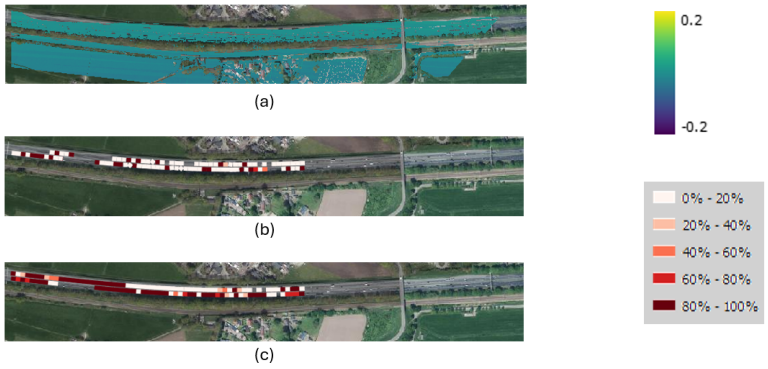


Figure 4.19: Test 4 benchmark distribution. (a) DTM residual map. (b) Automatically extracted benchmarks. (c) Manually digitised benchmarks

Figure 4.18 and Figure 4.19 present the benchmark distribution and residual DTM maps for Test 3 and Test 4. In Test 3 (Figure 4.18), both the automatically extracted and manually digitised road markings are mostly located on the northern lane. This uneven distribution appears to result in a gradient pattern in the DTM residual map, with greener tones in the north and more yellowish tones in the south, indicating alignment differences across the road width. In contrast, Test 4 (Figure 4.19) shows a more consistent green colour in the DTM residual map. Although the automatic extraction in Test 4 includes many unstable road markings, the stable ones are better distributed across both lanes. This suggests that having well-distributed, high-stability road markings plays a more important role in achieving uniform alignment than simply having a high number of extracted features.

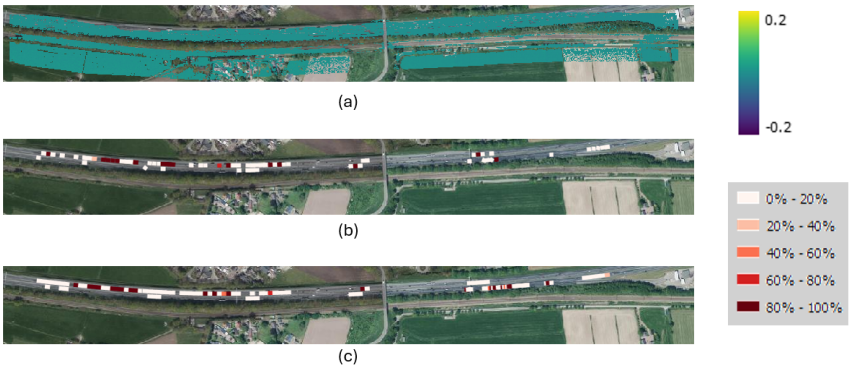


Figure 4.20: Test 5 benchmark distribution. (a) DTM residual map. (b) Automatically extracted benchmarks. (c) Manually digitised benchmarks

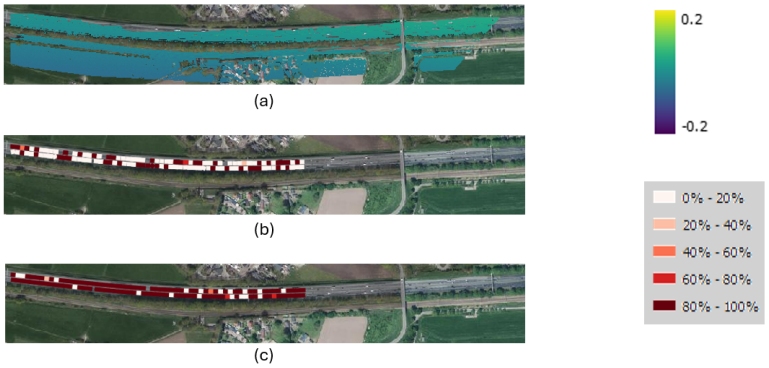


Figure 4.21: Test 6 benchmark distribution. (a) DTM residual map. (b) Automatically extracted benchmarks. (c) Manually digitised benchmarks

In Test 5, both the automatically extracted and manually digitised road markings (Figures 4.20b and 4.20c) are not well distributed, with most features concentrated on the northern lane and very few in the southern lane. Despite this, the DTM residual map (Figure 4.20a) shows a consistent green colour across the area, indicating a uniform alignment result. In contrast, Test 6 presents a more unexpected outcome. The road markings from both sources (Figures 4.21b and 4.21c) are well distributed across both lanes and appear more complete, yet the DTM residual map (Figure 4.21a) shows a slight gradient from north to south. This finding is surprising, as better benchmark distribution in Test 6 does not lead to a more consistent residual result compared to Test 5. It suggests that even when the input features appear well balanced, other factors in the alignment process might still influence the outcome. Therefore, visual inspection remains important alongside statistical evaluation, as it helps reveal inconsistencies that might not be obvious from metrics alone.

The results demonstrate that errors in automatically extracted road markings, such as incorrect segmentation, positional shifts, and inclusion of features that are not actual road markings, can influence the alignment outcome, particularly by affecting the distribution and stability of available benchmarks. However, the relationship between extraction quality and alignment accuracy is not straightforward. As shown across the test cases, high

completeness or correctness scores do not always correspond with better alignment results. Similarly, cases with balanced and well-distributed benchmarks sometimes still produce misalignments. These findings indicate that while extraction errors contribute to alignment inaccuracies, their exact impact is difficult to quantify using the proposed approach in this research. This is because the inaccuracies in the extracted road markings may differ between the reference and target datasets, making the remaining misalignment a combined effect of errors from both sides and also the quality of spatial distribution of the benchmark, which remains unexplored through the approach in this research. This highlights a limitation of the approach proposed in this research, which lacks a controlled way to trace and quantify how much the extraction errors contribute to the resulting alignment errors. The challenge is mainly due to the role of the spatial distribution of the benchmark, which acts as the link between extraction quality and alignment quality. However, a method to quantify the quality of this distribution has not been explored in this research.

To answer the third sub-research question: the accuracy of automatically extracted road markings influences LiDAR co-registration by affecting the distribution of the benchmarks used for alignment. Errors in the extraction process propagate to the quality of the benchmark distribution. However, the impact of errors in the reference and target datasets on the quality of this distribution remains unclear, as the current approach does not quantify the quality of the benchmark distribution.

4.4 Alignment Accuracy at Varying Distances from Road Markings

The section evaluates how alignment accuracy changes with increasing distance from the benchmark area. In Section 4.3, it was shown that the spatial distribution of road markings, such as when features are only present on a single lane or represented on both lanes of the highway, affects alignment accuracy. Furthermore, findings from both Section 4.3 and Section 4.2 indicate that areas located farther away from the benchmark tend to experience greater alignment errors. Building on these observations, this section investigates how alignment performance degrades over distance by dividing the DTM residual map into multiple segments extending away from the road marking benchmarks as proposed in Section 3.9.3. The average residual value is calculated for each segment, allowing a comparison of how alignment accuracy varies spatially. This analysis is carried out for both manually digitised and automatically extracted road markings to understand how alignment errors propagate beyond the benchmark area.

As mentioned in Section 3.9.3, this evaluation investigates how alignment accuracy changes with increasing distance from the area where road markings are used as benchmarks. To apply the approach, the alignment is performed by registering AHN3 to AHN5 using road markings located on a highway in the southern part of the extended AOI. These road markings serve as the only control features used for the transformation, allowing the remaining area, extending northward from the benchmark location, to be evaluated for changes in alignment accuracy over distance. The results of the alignment error per segment are shown in the Table 4.13.

Table 4.13: Mean DTM difference per segment for the alignment of AHN3 to AHN5

Mean DTM Difference per Segment (m)					
Segments	Segment 0	Segment 1	Segment 2	Segment 3	Segment 4
Untransformed	-0.033	-0.030	-0.020	-0.020	-0.027
Manual	0.001	-0.006	-0.017	-0.015	-0.022
Automatic	-0.019	0.147	0.318	0.486	0.645

The results in Table 4.13 show a noticeable increase in alignment error for automatically extracted road markings as the distance from the benchmark increases, while the manually digitised benchmarks maintain low and stable DTM difference values across all segments. When using manually digitised benchmarks, the DTM difference remains low and stable across all segments. The results show that the alignment remains consistent even for segments that were far from the area where the road markings are. In contrast, the automatic extraction shows a significant increase in DTM difference with distance. The results indicate that the alignment accuracy degrades as the segment increases. Overall, the results show that manually digitised road markings are better at maintaining alignment accuracy as the distance from the benchmark area increases. In contrast, automatically extracted road markings lead to larger alignment errors in the farther segments, while in Segment 0, where the road markings are located, the error from automatic extraction remains relatively small (around 2 cm). The result suggests that the impact of extraction errors becomes more significant further from the benchmark area.

Table 4.14: Difference between adjacent segment of AHN3 to AHN5 alignment

Difference Between Adjacent Segments (m)		
Extraction Method	Manual	Automatic
Δ between Segment 0 and 1	-0.007	0.166
Δ between Segment 1 and 2	-0.011	0.170
Δ between Segment 2 and 3	0.002	0.169
Δ between Segment 3 and 4	-0.007	0.159
Mean Δ	-0.006	0.166

Table 4.14 presents the differences in mean DTM residuals between adjacent segments to further support the findings. The results show that manually digitised road markings result in smaller differences between segments compared to automatically extracted ones. For the manual method, the changes in residual values are minimal, within a few millimetres. In contrast, the automatic method shows significantly larger and more consistent increases in residual differences between segments, with values around 16 cm. The result shows that alignment accuracy using automatically extracted road markings tends to degrade faster with distance than the manually digitised road markings. To better understand the source of this difference, visual inspection is performed, with the results presented in the following figures.

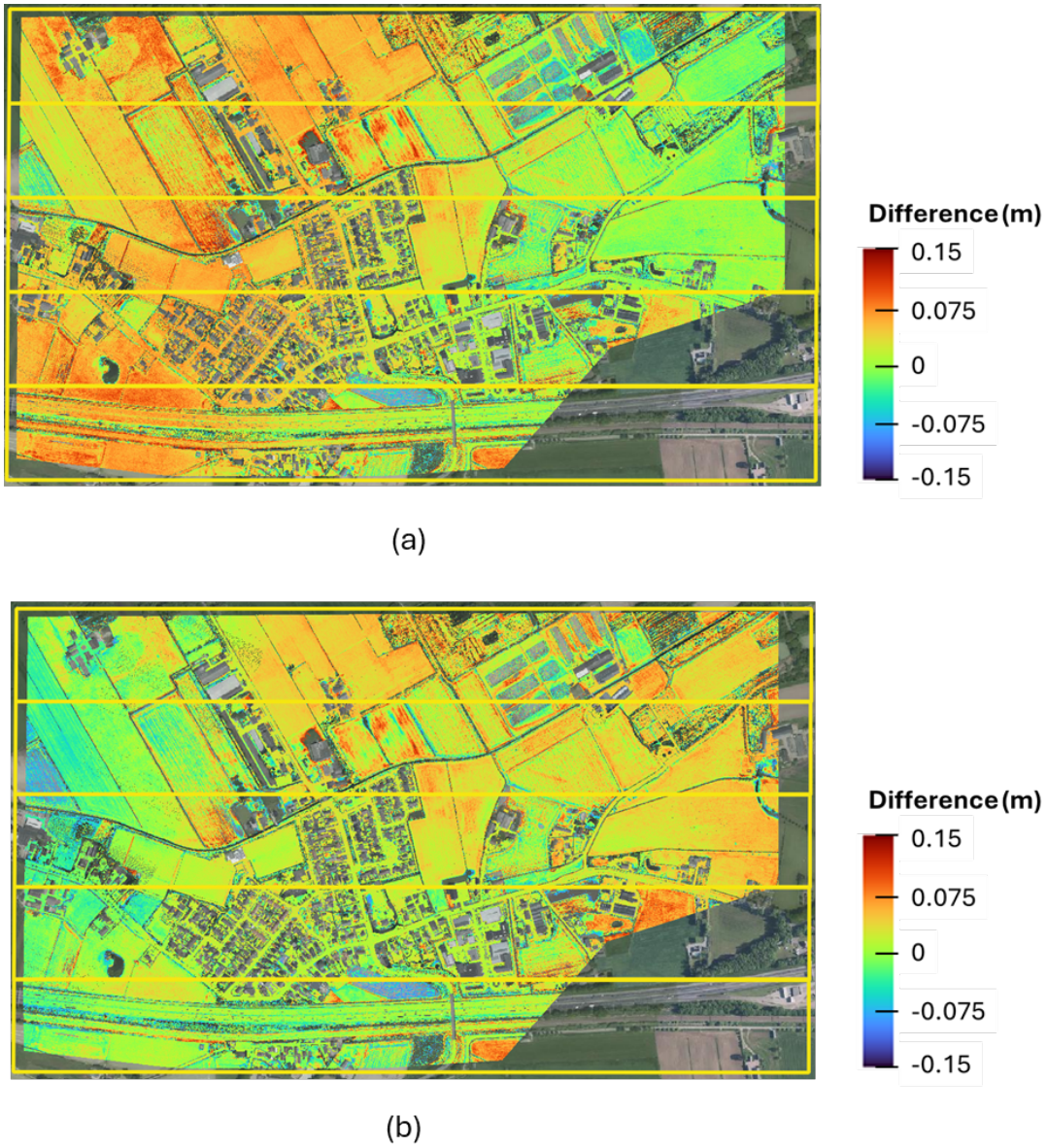


Figure 4.22: DTM difference map for alignment of AHN3 to AHN5 using manually digitised road marks. (a) Untransformed. (b) Aligned using manually digitised road markings.

Figure 4.22 presents the comparison of DTM difference maps for the alignment of AHN3 to AHN5 using manually digitised road markings. In the untransformed case (Figure 4.22a), red and orange tones dominate, especially in the left part of the map. After alignment using manually digitised road markings (Figure 4.22b), the colours shift closer to green and yellow, which represent values near zero, indicating improved alignment. Although there are still some red and orange areas in the right part of the map, they appear lighter and more scattered, showing that alignment accuracy has improved, particularly in the area where the benchmark is located, the polygon in the southernmost part of the map.

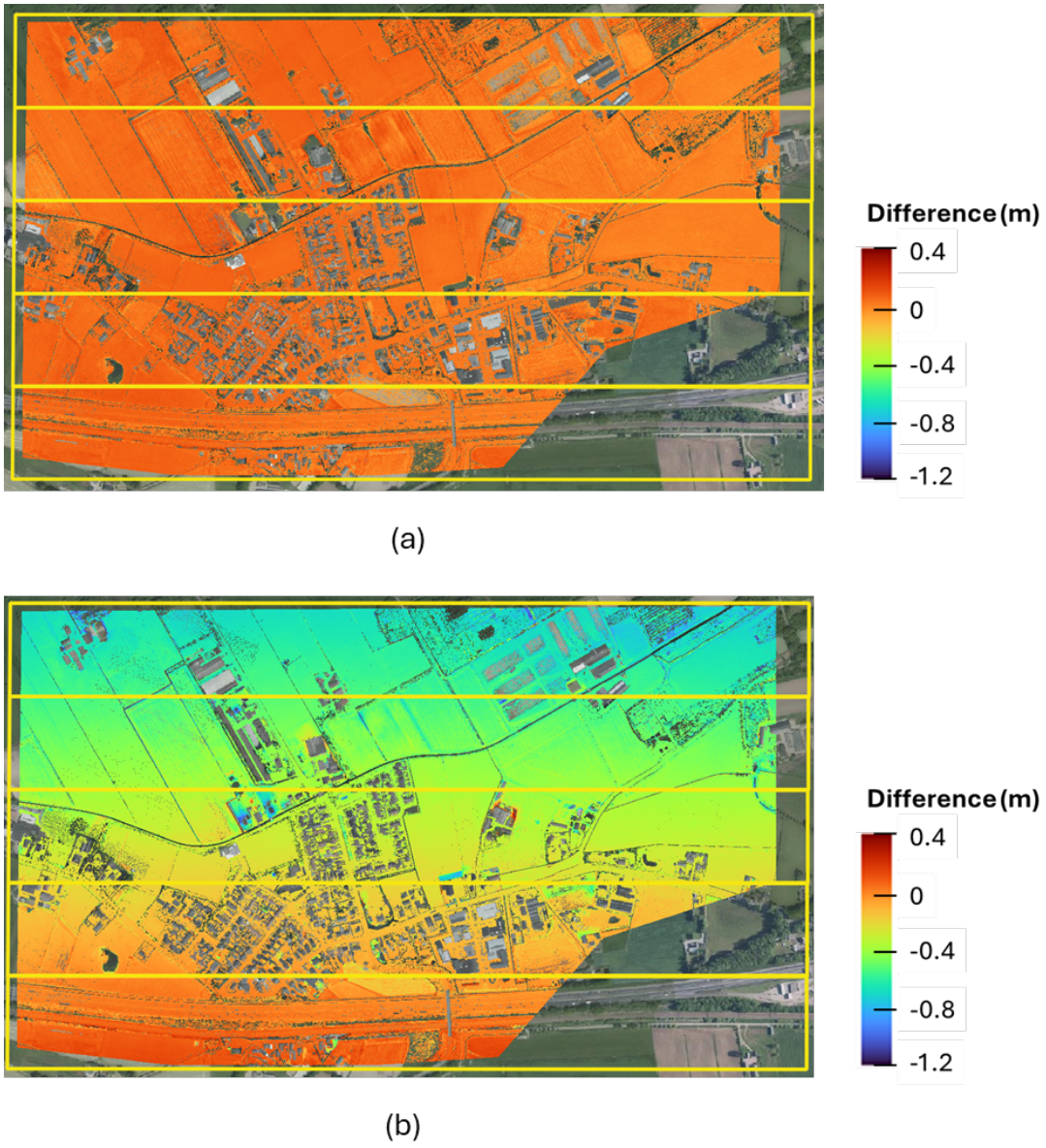


Figure 4.23: DTM difference map for alignment of AHN3 to AHN5 using automatically digitised road marks. (a) Untransformed. (b) Aligned using automatically digitised road markings.

Figure 4.23 presents the DTM difference maps for the alignment of AHN3 to AHN5 using automatically extracted road markings. In the untransformed case (Figure 4.23a), most of the map is covered in red and orange tones, indicating consistent elevation offsets across the area. After alignment using automatically extracted road markings (Figure 4.23b), the colour pattern becomes more varied, with blue and green tones dominating the northernmost segments and red and orange tones dominating the southernmost segments. This variation in colour of the heat map indicates alignment errors introduced in some areas. While the

southern segment, where most road marking benchmarks are located, shows relatively low alignment error, the northern segments display significant misalignments, meaning that the accuracy of alignment increases with distance from the control features. To better understand why the DTM difference maps from manually digitised and automatically extracted road markings differ significantly, a visual inspection of the road markings used in the alignment will be conducted.

Figure 4.24: Distribution of benchmark road markings within the validation AOIs. (a) Manually digitised road markings. (b) Automatically extracted road markings.

Figure 4.24 shows the spatial distribution of benchmark road markings within the validation AOI. It can be seen that road markings are only located in the southernmost segment of the AOI. Furthermore, due to the limited spatial coverage of the AHN5 dataset, only the left half of Segment 0 is fully covered. A more detailed view of the road marking distribution is provided in the following figure.

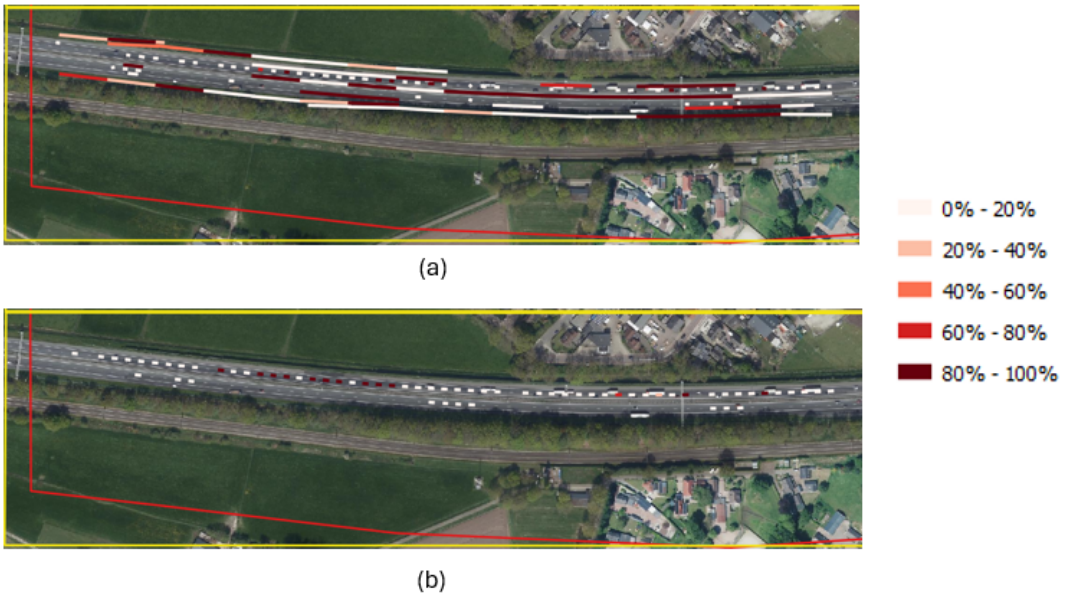


Figure 4.25: Closer view of road marking distribution of alignment AHN3 to AHN5. (a) Manually digitised road markings. (b) Automatically extracted road markings.

The visual comparison in Figure 4.25 reveals a key difference between the automatically extracted and manually digitised road markings in the benchmark area. In the automatic case (Figure 4.25a), several features appear along the road edges that are not present in the manual dataset (Figure 4.25b). These are non-road marking elements, such as road edge lines, that were mistakenly included during the automatic extraction process. Since these features were not successfully filtered out during either the extraction or the correspondence establishment step, they may have been wrongly used as benchmarks. This introduces false correspondences, which might distort the transformation estimation and negatively impact the alignment accuracy. Based on this finding, improvements are recommended in both the extraction process and the correspondence establishment step to better filter out non-road marking objects for future work.

To better understand the pattern observed in the DTM difference results, the rotation parameters derived from the transformation matrices are compared. These parameters indicate how the target point cloud was rotated to align with the reference. In this context, the rotations are expressed as Omega (rotation around the X-axis), Phi (rotation around the Y-axis), and Kappa (rotation around the Z-axis). The consistent differences between adjacent segments shown in Table 4.14 suggest that there is a notable difference in rotation around the X-axis between the alignment results. Table 4.15 summarises the rotation angle differences between the manual and automatic alignment results for the AHN3 to AHN5 case.

Table 4.15: Comparison of rotation angles (Omega, Phi, Kappa) between manual and automatic road marking alignment for AHN3 to AHN5.

Rotation Axis	Manual RM	Automatic RM	Difference
Omega (°)	0.000740	-0.046187	-0.046927
Phi (°)	-0.004176	-0.000962	0.003214
Kappa (°)	0.005920	0.004096	-0.001824

After comparing the transformation matrices in Table 4.15, it becomes clear that the largest difference occurs in the Omega angle, which represents rotation around the X-axis. This is likely because the road that contains the extracted road markings naturally runs along the X direction. As a result, the road markings are well distributed along the X-axis but lack sufficient distribution along the Y-axis, making the alignment more sensitive to rotation errors in that direction.

To conclude, this section shows that alignment accuracy tends to decrease as the distance from the benchmark area increases, particularly when using automatically extracted road markings. Manually digitised road markings maintain low and stable alignment errors across all segments, while automatically extracted road markings show a consistent increase in DTM residuals. This increase in error is also reflected in the difference between adjacent segments, where automatic methods show large and consistent jumps in error, while manual methods show small, consistent jumps. Visual inspection reveals that false features, such as road edge lines included during automatic extraction, affect the difference in rotation along the X-axis. However, there is a limitation in this approach: it remains unclear how differences in road marking distribution between manual and automatic methods directly affect the rotation outcome.

To answer the fourth sub-research question: alignment error increases consistently with distance from the area where road markings are used as control features. This is because the error originates from a rotation matrix that does not accurately capture the correct rotation during the alignment process, leading to growing misalignments further from the control area. However, the factors that determine the extent of this increase cannot be quantified using the approach applied in this research.

4.5 Type of Uncertainty in Position Harmonisation Using Automatic Extracted Road Markings

This section identifies and categorises the uncertainties that influence the accuracy of position harmonisation when using automatically extracted road markings as reference features. These uncertainties are examined by reviewing the proposed methodology and the evaluation results obtained in this research. Uncertainty in the alignment process refers to factors that reduce confidence in the accuracy of the result. These factors can occur at different steps in the alignment process and affect the following steps, leading to errors in the final alignment. The uncertainties presented in this section were identified during the research process through analysis of each alignment stage and evaluation of the resulting outputs. In this section, the identified uncertainties are organised according to the steps of the alignment process, as illustrated in Figure 4.26.

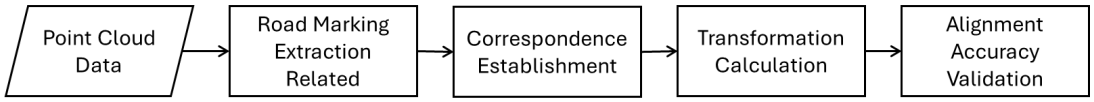


Figure 4.26: Alignment process pipeline

4.5.1 Data-Related Uncertainties

The first type of uncertainty originates from the characteristics of the input point cloud data. These uncertainties are not introduced by processing steps but are inherent to the datasets themselves. In this research, the reference and target point clouds were acquired from different projects, resulting in variations in point density and acquisition time. These differences introduce uncertainty at the very beginning of the alignment process, which propagates through the alignment steps and affects the final alignment result. The uncertainties included in this category are described below.

Variation of Point Cloud Density

To begin with, the point cloud density and intensity range of road markings in each dataset were calculated in Section 3.2.4 to understand the characteristics of the input data. The results are presented in Table 3.1. Further analysis was conducted in Section 4.3.1, where the accuracy of automatically extracted road marking lines was evaluated. The results shown in Table 4.9 indicate that the Prorail datasets produced a lower mean 3D centroid difference compared to the AHN datasets. This performance is related to the significantly higher point

density of the Prorail datasets compared with AHN datasets shown in Table 3.1. These findings demonstrate that point cloud density has a direct impact on the accuracy of line extraction, particularly the 3D centroid position, which in turn affects the stability of correspondences in the correspondence establishment process

Variation of Point Cloud Acquisition Time

Based on the findings in Section 4.1, the number of candidate correspondences is related to the time gap between point cloud acquisitions. Smaller time gaps result in a higher number of candidate road markings that proceed to the next step of correspondence establishment. A greater number of correspondences reduces the risk of over-filtering in the RANSAC-weighted alignment method, as discussed in Section 4.2. This also increases the likelihood of achieving a more balanced spatial distribution of road markings, particularly between the northern and southern lanes of the highway, which contributes to improved alignment accuracy, as analysed in Sections 4.2 and 4.3.

4.5.2 Extraction-Related Uncertainties

The next type of uncertainty comes from the road marking extraction process. This process was carried out using an adaptive intensity filtering approach to handle variations in the intensity range of the input data as introduced in Heide 2024. As described in Section 2.2, the method consists of multiple steps to produce line features that serve as benchmarks for the co-registration process (see Table 2.3). However, as discussed in Section 4.3.1, the extracted lines are affected by extraction errors, which in turn influence the alignment accuracy. Inspection of several extracted road marking samples revealed uncertainties in the extraction results that may propagate through the alignment process and affect the final accuracy.

Road Marking Clustering Uncertainty

Road Marking Cluster filtering is applied to remove high-intensity objects other than road marking clusters by evaluating each cluster's geometric features against road marking geometry threshold values as shown in Table 2.4. However, as shown in Figure 4.8, some edge lines were still included as the road markings to be used as benchmarks in the co-registration process. This indicates that the geometric filtering step is not always sufficient to distinguish between different types of high-intensity features, particularly when non-road features share similar geometric properties with actual road markings.

Line Segmentation Uncertainty

Line segmentation is performed using principal component analysis and convex hull fitting to convert point clusters into line features, as reviewed in Section 2.2, to fit a line into the road marking segments. However, as shown in Figure 4.14 and 4.15, some extracted lines' shapes did not represent the road marking very well, while another sample shows that the line is shifted from its road marking points, leading to an error in centroid accuracy.

Zigzag and Direction Correction Uncertainty

Zigzag filtering and direction correction are applied to remove line segments that deviate from the expected road orientation, based on comparison with road direction data obtained from the Dutch National Road Database (NWB). However, as shown in Figure 4.9 and Figure 4.10, some features with orientations nearly perpendicular to the road were still included as benchmarks for the co-registration process.

4.5.3 Correspondence Establishment-Related Uncertainties

Another type of uncertainty comes from the proposed correspondence establishment method described in Section 3.6. The method depends on the centroid of each line feature as the key element for establishing correspondences, which may introduce uncertainty due to potential errors in centroid estimation. In addition, the method uses distance thresholds in both the candidate correspondence selection step and the inlier evaluation step. The choice of these threshold values can affect the correspondence results.

Centroid Estimation Uncertainty

The uncertainty comes from the use of 3D centroids to represent road marking segments during the correspondence process. When a line is inaccurately extracted, such as when the shape of the line did not accurately represent the road marking (as shown in Figure 4.14) or the position of the line is shifted from its original position (shown in Figure 4.15), the calculated centroid may not reflect the true position of the actual road marking. Such centroid shifts can affect the stability of a correspondence and influence the result of the transformation calculation.

Distance Threshold Uncertainty

The correspondence method relies on spatial distance thresholds to determine which features are considered candidate matches and inliers. These thresholds directly affect how many correspondences are included or rejected. During the inlier evaluation, if the threshold is too small, the number of road markings identified as highly stable will decrease. Conversely, if the threshold is too large, more road markings will be classified as stable. This influences the transformation calculation process, as the stability of each road marking is used as a weight in the computation.

4.5.4 Transformation Calculation-Related Uncertainties

Uncertainties may also occur during the calculation of the transformation matrix used to align the target dataset to the reference. In this research, the transformation is computed based on correspondences established using RANSAC inlier voting, where the number of times each correspondence is selected as an inlier is used as a weight, as described in Section 3.7. However, using the inlier count as a weight causes correspondences with very low inlier counts to be excluded from the transformation calculation. This changes the spatial distribution of the benchmarks used in the transformation process.

Spatial Distribution of Benchmark

Uncertainty in the transformation calculation can arise from the spatial distribution of the correspondences used in the transformation process. As analysed in Section 4.2, where the transformation results using the RANSAC-weighted alignment method are compared with the nearest neighbour search method, a significant difference in alignment accuracy is observed for the case of aligning Prorail 2021 to AHN3. Figure 4.1 shows that the stable benchmarks identified by the RANSAC-weighted method are concentrated only on the north lane of the highway, with no stable benchmarks found on the south lane. This uneven distribution leads to different alignment results between the two methods. The nearest neighbour method, which retains some benchmarks in the south lane, produces better alignment accuracy. This demonstrates that the spatial distribution of benchmarks plays an important role in determining the accuracy of the transformation. A similar condition is observed in Sections 4.3 and 4.4, where differences in the spatial distribution of benchmarks, resulting from the correspondence establishment process using manually digitised versus automatically extracted road markings, lead to variations in alignment accuracy.

4.5.5 Alignment Accuracy Validation-Related Uncertainties

The final stage of the alignment pipeline involves evaluating the accuracy of the alignment results using a DTM differencing method as described in Section 3.8. This process compares ground surface elevations between the reference and target point clouds, assuming minimal environmental change. While this method enables large-scale quantitative validation, it also introduces uncertainties due to specific processing steps such as ground filtering, raster resampling, and outlier removal. These steps may influence the final DTM difference values and, consequently, the interpretation of alignment accuracy.

Ground Filtering Uncertainty

The DTM differencing relies on filtering the point clouds to retain only ground points. However, in datasets without ground classification labels, such as Prorail datasets, this step is performed using an automated filtering algorithm, such as CSF in CloudCompare, which is implemented in this research. This introduces uncertainty, especially in areas with low vegetation such as crops or bushes, which may be misclassified as ground, leading to errors in the DTM differencing process.

Rasterisation Uncertainty

During rasterisation, the elevation of each grid cell is calculated by averaging all points inside the grid. This averaging process can make small changes in elevation less noticeable, which makes it more difficult to detect small vertical alignment errors. Additionally, the choice of raster resolution may affect the accuracy of the comparison. A larger pixel resolution may fail to detect small alignment errors, while a smaller pixel resolution may leave some grid cells empty if there are no points inside the grid, especially in areas with low point density. These factors introduce uncertainty into the alignment accuracy.

Outlier Filtering Uncertainty

The research utilises IQR filtering on the DTM difference raster to remove noise on the ground surface caused by changes unrelated to alignment, such as vegetation or excavation. While this step reduces some of the noise, it is still difficult to determine whether there is any remaining noise that contributes to the evaluation results or not.

4.5.6 Summary of Identified Uncertainties

To answer the fifth sub-research question: the following section presents five main types of uncertainty that influence the accuracy of position harmonisation when using automatically extracted road markings. Each type corresponds to a step in the alignment process and was supported by findings from the methodological review and evaluation results. Table 4.16 provides an overview of these uncertainty types and summarises their impacts.

Table 4.16: Summary of identified uncertainties in the alignment pipeline

Type	Uncertainty	Observed Impact
Data-Related	Variation in point cloud density	Affects the accuracy of extracted road marking centroids
	Variation in point cloud acquisition time	Affects the number candidate correspondences
Extraction-Related	Road marking clustering uncertainty	Non-road marking features are classified as benchmark
	Line segmentation uncertainty	Shifts centroid position, affecting correspondences establishment
	Zigzag and direction correction uncertainty	Non-road marking features are classified as benchmark
Corr. Establishment-Related	Centroid estimation uncertainty	Affects number of correspondences and their stability
	Distance threshold uncertainty	Affects number of correspondences and their stability
Transformation Calculation-Related	Spatial distribution of benchmarks	May produce incorrect rotation and translation values of alignment
Acc. Validation-Related	Ground filtering uncertainty	Includes non-ground points, reducing DTM accuracy
	Rasterisation uncertainty	Choice of pixel size affect DTM differencing result
	Outlier filtering uncertainty	Unremoved noises affect DTM differencing result

5 Discussion

5.1 Selection of Study Case and Study Area

5.1.1 Limitation of Study Case

Although this research provides useful insights into the use of road markings as benchmarks for LiDAR point clouds co-registration, several limitations need to be considered when analysing the results. These limitations mainly relate to the characteristics of the study area and the datasets used, which may affect how broadly the findings can be applied to other environments or conditions.

First, the study focuses only on road markings located on highways. Highways usually have well-maintained markings, clear lane separations, and relatively consistent road surfaces. This makes them suitable for testing automatic extraction methods. However, other road types, such as urban streets, rural roads, or intersections, often have different surface materials or levels of fade or tear. These differences can affect the visibility and shape of the road markings in the point cloud data. Since these road types were not included in the study, the performance of the extraction method and alignment approach in those conditions remains unknown. Therefore, the method might not work equally well in more complex environments.

Second, the highway sections analysed in this study all run in a west-to-east or east-to-west direction and include only two lanes. This introduces a limitation in the evaluation of road marking distribution. In this study, only two situations are observed: one where road markings are spread evenly across both lanes, and one where the markings appear on only one side. However, in real-world applications, road markings can be more complex, such as at roundabouts, cross-overs, or areas with construction. The lack of variation in road structure means the study cannot fully explore how uneven or imbalanced marking distributions might impact the alignment results, especially for the proposed alignment method in this research.

Lastly, the datasets used in this study are limited to airborne LiDAR acquisitions, where Prorail datasets are collected from a helicopter and AHN datasets are collected from an aeroplane. While these datasets provide high-quality top-down views of road environments, they do not represent all possible acquisition methods. For example, MLS and TLS provide data from ground level and often have higher density and different noise characteristics. These methods may capture road markings differently and could influence both the feature extraction process and the quality of the final alignment. Because only airborne datasets are used in this study, the findings may not generalise well to point clouds captured using other techniques.

5.1.2 Limitation of Study Area

In a comparison study using different datasets, it is important that all datasets cover the same area. When datasets differ in terms of their availability or extent, they can unintentionally introduce bias into the results. This situation happened in this research due to differences in the release schedule of the AHN5 dataset.

At the time of this research, AHN5 data had not yet been fully released for the entire area of interest. As shown in Figure 5.1, only the blue region (Regio West 2023) had AHN5 coverage, while the red region (Regio Oost-Zuid 2025) was still unavailable. Because of this, the analysis using AHN5 was limited to a smaller area, while other datasets were tested on a larger, more complete region. This discrepancy in data coverage creates bias in the evaluation.

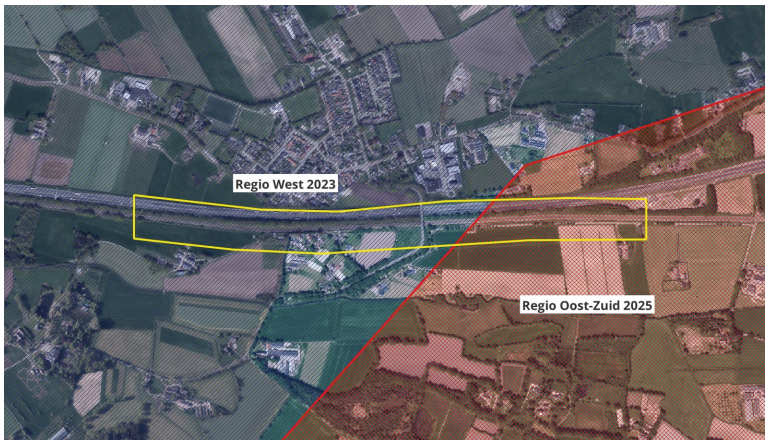


Figure 5.1: AHN5 data availability during the time of research. Only the blue area (Regio West 2023) had AHN5 data available when the research was conducted. The red area (Regio Oost-Zuid 2025) had not yet been released, limiting the AOI and potentially influencing detection performance.

From Table 4.8, the AHN5 dataset shows a perfect completeness score of 1.000, meaning all road markings in the tested area were successfully detected. However, this result should be interpreted with caution. It is possible that the unreleased area of AHN5 contains more difficult or lower-quality road markings, so the high completeness score might be due to the easier conditions in the available section. It is not clear if AHN5 would perform the same if tested on the full area.

In this study, the AOI was selected before checking which AHN5 tiles were available. For future research, it may be better to first review the available data and then define the AOI accordingly. This would help avoid issues caused by differences in data coverage and ensure a more balanced comparison between datasets

5.2 Determination of Z-Score Threshold to Filter Low Intensity Points in Manually Segmented Road Marking

In this research, a z-score threshold was used to filter out low-intensity points from manually segmented road marking clusters. This was done to improve the accuracy of geometry feature calculations by removing noise from non-road marking points, such as pavement surfaces that were included in the manual segmentation process. The z-score filtering process depends on the intensity values within each segmented cluster, where the threshold helps to determine points that have significantly lower intensity values than the actual road marking. However, the effectiveness of the z-score filtering is highly dependent on the consistency of the manually created segmentation polygons. One limitation of this manual segmentation approach is that it is difficult to manually draw polygons of the same size and shape for each road marking. Since the segmentation process was done visually in CloudCompare, the size of the selected area can vary depending on the operator, especially when the road markings are visually unclear.

This inconsistency directly affects the average intensity value calculated for each road marking segment. If a larger polygon is used, it is more likely to include additional low-intensity points from surrounding pavement, while the number of true high-intensity road marking points remains the same. As a result, the average intensity for that segment becomes lower, shifting the z-score values of the points. This causes the z-score threshold to perform differently across segments, depending on the size and content of the manually drawn polygon. In practice, this means that the same threshold may filter different amounts of points for each segment, depending on how the segmentation was done. For some segments, the threshold may remove only the unwanted low-intensity points as intended. For others, it may retain too many or filter out useful points because the average intensity was shifted due to inconsistent polygon sizes. Therefore, while the z-score method is useful for cleaning segmented road markings, its performance is affected by the manual segmentation step, making it less reliable when road markings are manually segmented.

Figure 5.2 illustrates how the size of the segmentation polygon affects the result of z-score thresholding on road marking segments. The top row shows examples where the polygon is drawn tightly around the road marking, capturing mostly high-intensity points. As a result, the z-score calculation retains only a small number of points, which are concentrated along the marking. In contrast, the bottom row shows results when a larger polygon is used, which includes more surrounding pavement points with lower intensity. This lowers the average intensity value for the segment, causing more points to pass the z-score threshold. Assuming that the orthophoto basemap provides an accurate reference for the actual size of the road marking, the optimal z-score threshold for the top row segment (RM_21) is approximately -1.5, as it preserves the highest point density within the marking area. In contrast, for the bottom row segment (RM_35), a threshold of -0.5 is more suitable, as it retains the points located within the marking and removes most of the pavement points included due to the larger segmentation area. This shows that the size of the segmentation process affects the z-score threshold value.

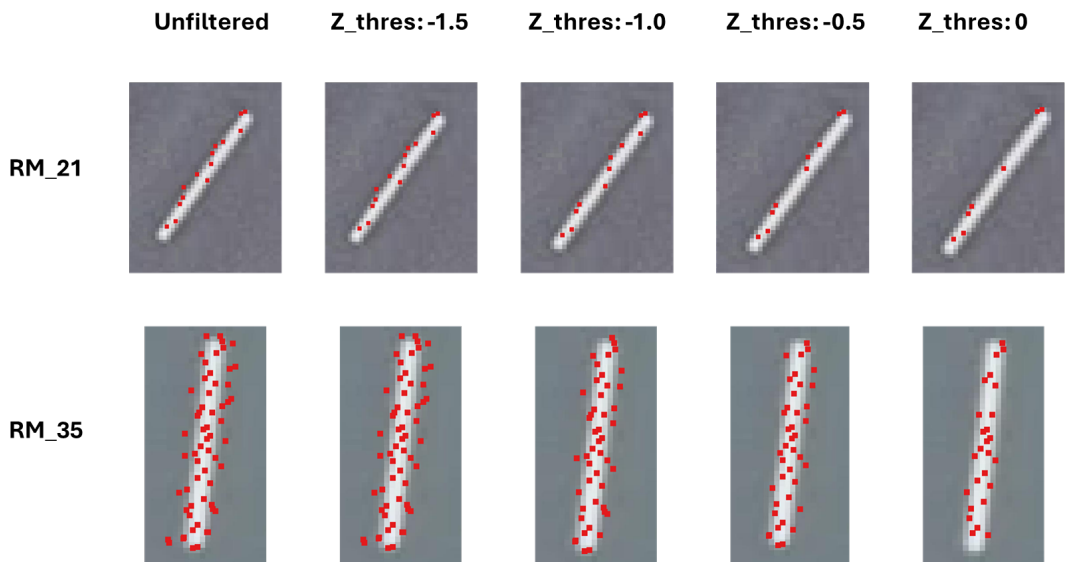


Figure 5.2: Illustration of z-score thresholding results on road markings using different polygon sizes. Top row: smaller segmentation polygon. Bottom row: larger segmentation polygon.

These findings highlight the need for an adaptive z-score threshold that can adjust to inconsistencies introduced during the manual segmentation process. Since operators may draw segmentation polygons with varying sizes or shapes, the average intensity value of each road marking segment can differ significantly. This affects the calculation of the z-score and results in inconsistent filtering performance across segments. An adaptive approach could help ensure that points are evaluated relative to the specific characteristics of each segment, rather than using a single fixed threshold for all.

5.3 Calculation of Geometric Features Threshold for Road Marking Segmentation

The threshold calculation was based on manually segmented road markings, following the workflow described in Section 3.3. For each road marking segment, six geometric features were calculated based on the eigenvalues of the point distribution. These features and their formulas are shown in Table 2.4. To better understand the range of each feature, histograms were created to show the distribution of values across all road marking samples. These histograms helped identify the most common value ranges and detect any outliers. To remove these outliers, the IQR filtering method was applied to each histogram. Based on this filtering, the minimum and maximum threshold values were estimated from the cleaned data distribution. The histograms of all geometric features before and after IQR filtering are shown below.

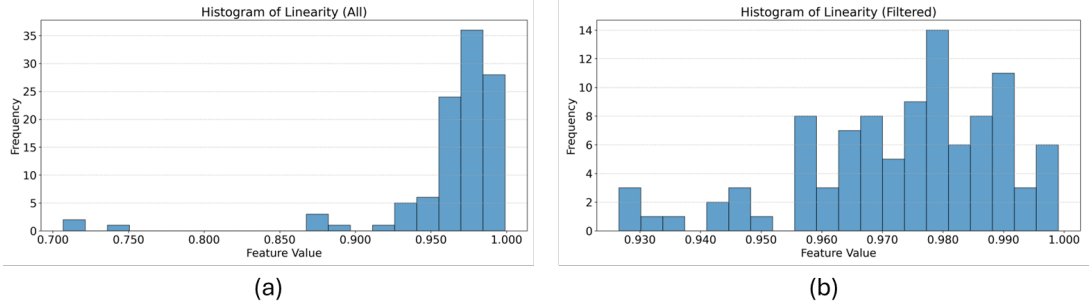


Figure 5.3: Histogram comparison of linearity feature values. (a) Before and (b) after IQR filtering

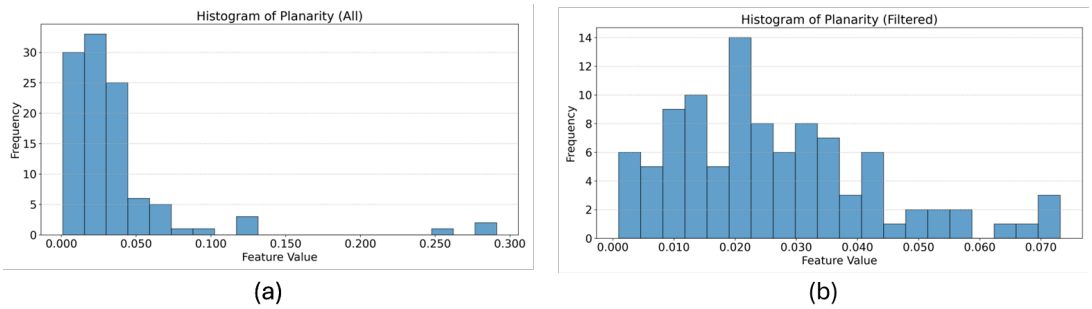


Figure 5.4: Histogram comparison of planarity feature values. (a) Before and (b) after IQR filtering

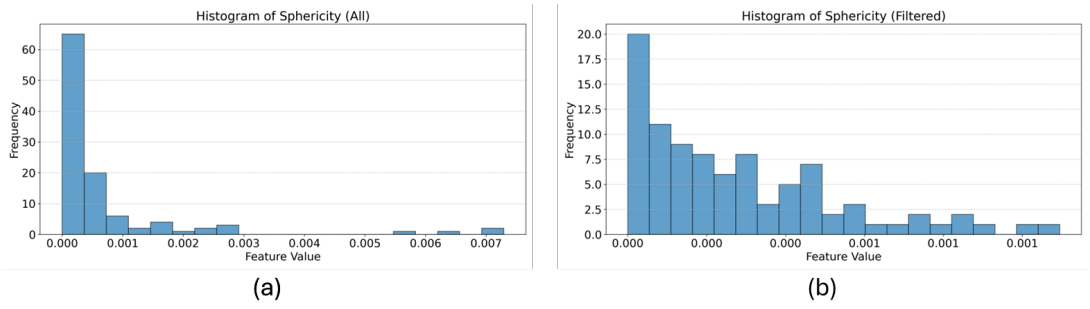


Figure 5.5: Histogram comparison of sphericity feature values. (a) Before and (b) after IQR filtering

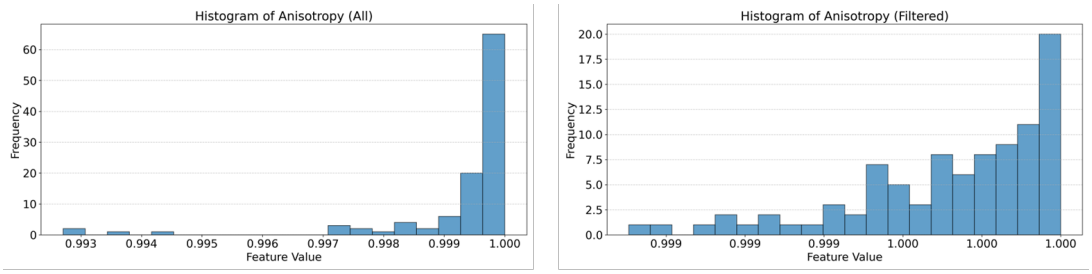


Figure 5.6: Histogram comparison of anisotropy feature values. (a) Before and (b) after IQR filtering

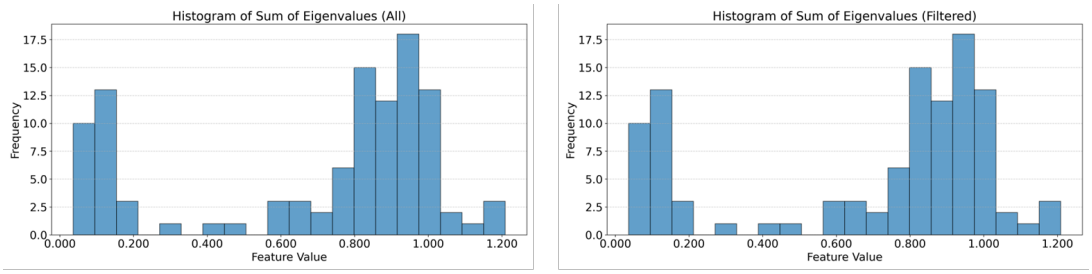


Figure 5.7: Histogram comparison of sum of eigenvalues feature values. (a) Before and (b) after IQR filtering

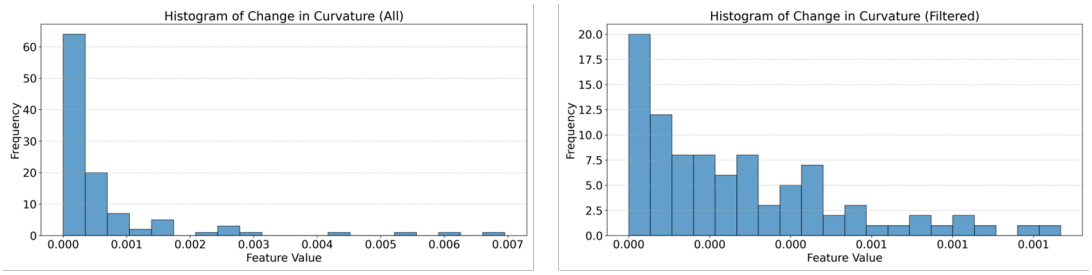


Figure 5.8: Histogram comparison of change in curvature feature values. (a) Before and (b) after IQR filtering

As shown in the histogram plots, most of the geometric features have a single peak in their distribution. This means that most of the road marking segments have feature values that cluster around a typical range. These types of distributions are suited for IQR filtering, which works by keeping only the values that fall within the central range of the data, between the first and third quartile (Dekking et al. 2005), and removing values that are much lower or higher than average, which are considered as outliers.

However, the sum of the eigenvalues features behaves differently. Its histogram shows two separate peaks instead of just one. Because of this, the IQR method does not work well. The IQR filter works by calculating the first quartile (Q1) and third quartile (Q3) of the data, and then only keeping values that fall between them, typically extended slightly using the $1.5 \times$ IQR rule Dekking et al. 2005.

But when the data has two different groups, this approach becomes ineffective, since the Q2 will fall between the two peaks. In the case of the sum of eigenvalues, the actual quartile values are $Q1 = 0.3360$ and $Q3 = 0.9413$, giving an interquartile range:

$$\text{IQR} = Q3 - Q1 = 0.9413 - 0.3360 = 0.6053 \tag{5.1}$$

The IQR filter then calculates the filtering bounds as:

$$\text{Lower bound} = Q1 - 1.5 \times \text{IQR} = 0.3360 - 1.5 \times 0.6053 = -0.5720 \tag{5.2}$$

$$\text{Upper bound} = Q3 + 1.5 \times \text{IQR} = 0.9413 + 1.5 \times 0.6053 = 1.8492 \tag{5.3}$$

This results in a very wide range that includes almost all values in the histogram. As a result, the IQR filter fails to remove any values. While the filtering formula works correctly, it is not suitable when the data has more than one peak. The wide IQR range makes it unable to remove noise values.

The sum of eigenvalues describes the overall size or spread of a point cloud segment. It is calculated by adding up the three eigenvalues of the covariance matrix, which tell us how much the points vary in space. A higher value means the segment is more spread out (either longer, wider, or denser), while a lower value means the segment is more compact or narrow.

In this case, the histogram of the sum of eigenvalues shows two peaks, which suggests that there are two different groups of road markings in the data. This makes sense, because during manual road marking segmentation, the dataset included mainly two different types of road markings, which are; short dashed lines and long dashed lines. These two types naturally have different spatial sizes, resulting two peaks in the sum of the eigenvalues histogram.

(a) (b)

Figure 5.9: Two types of road markings segmented for training sample. (a) Short and (b) long dashed road markings

Based on these findings, it is recommended that the geometric feature thresholds used to filter road marking segments in the automatic extraction method should be adjusted according to the type of road marking. The presence of two distinct peaks in the histogram of the sum of eigenvalues suggests that different types of markings, such as short dashed lines and long dashed lines, have different spatial characteristics. Differentiating thresholds by marking type could improve the accuracy of the segmentation. This observation also opens up a direction for future research, not only for improving road marking extraction, but also for developing methods to classify different types of extracted road markings based on their geometric features.

The final minimum and maximum threshold values for each geometric feature, as determined through IQR filtering, are summarised in Table 3.2. These values were extracted from the cleaned histograms of each feature after removing outliers. They represent the range of geometric features observed in valid road marking segments. These thresholds are intended to support the automatic road marking extraction method by helping to distinguish road markings from other objects in the point cloud.

5.4 Limitations in Alignment Evaluation and Interpretation

5.4.1 Challenges in Visual Inspection for Analysing Error Sources

In this research, visual inspection was used to assess both the detection and geometric accuracy of automatically extracted road markings. For detection accuracy, the method involved overlaying the extracted and reference road markings on ground-classified point clouds, with intensity values adjusted to make road markings more visible. This was effective in identifying major detection errors, such as missing markings (false negatives) or incorrectly extracted features (false positives). However, while these errors could be observed, their underlying causes could not be fully explained through visual inspection alone.

For example, as shown in Figures 4.5 and 4.6, some undetected road markings still exhibit high intensity, valid geometric features, and no elevation change that would indicate removal due to curb filtering. This suggests that the error did not occur in the early stages, such as intensity filtering, geometry thresholding, or elevation filtering. Instead, the problem may have originated from later steps like road marking clustering, line segmentation, or zigzag and direction correction. Since these steps cannot be directly interpreted from the point cloud, the explanation of the error remains an assumption.

Similarly, for line accuracy, 2D visual inspection was used to verify whether the extracted lines were correctly placed and oriented. This method helped detect horizontal misalignments and direction errors. However, it did not provide insight into why these errors occurred. This is because visual inspection only reveals the final output of the pipeline, not the intermediate results, such as the result of adaptive intensity filtering in removing low-intensity points, the result of clustering high-intensity points, or which lines were removed during the zigzag and direction correction step. Therefore, although visual analysis is useful for detecting errors, it is limited in its ability to trace their exact origin within the extraction process.

5.4.2 Difficulty in Understanding Transformation Instability

The results of this research show that the spatial distribution of road markings has an influence on the quality of the alignment, as identified in Section 4.2 and 4.3. The alignment accuracy decreased in areas that were far from where the road markings were concentrated, as shown in Section 4.4. This issue is linked to differences in the transformation matrix used during alignment, especially in the rotation around the x-axis. This makes sense because the road markings were mostly distributed along the x-direction of the road, meaning the control points were unevenly spread in space. However, the current approach does not include a way to measure or quantify how well the benchmarks are distributed. Because of this, it is difficult to explain exactly how the quality of the distribution relates to the alignment errors. Future research is needed to better understand this relationship, especially how the density and spread of road markings in different directions affect the rotation and translation used in point cloud alignment.

6 Conclusion and Recommendations

6.1 Conclusion

As proposed in Section 1.2, this research is conducted to demonstrate the use of road marking extracted using the method proposed in Heide 2024 as a benchmark in the LiDAR point cloud harmonisation process to understand how its extraction accuracy affects the final alignment accuracy. Following that, a main research question is proposed to reflect the objective of the research:

"How suitable are road markings extracted using the method proposed in Heide 2024 as a benchmark for heterogeneous LiDAR point cloud harmonisation?"

In this context, suitability refers to how well the extracted road markings fulfil their role as a benchmark for LiDAR harmonisation. This involves assessing whether they consistently reduce misalignment across datasets acquired at different times and with different sensors, despite the presence of inaccuracies in the extraction process

This takes into account that datasets acquired at different times may be affected by environmental changes, while different sensors introduce variations in dataset specifications. These factors may influence the availability and consistency of road markings as benchmarks between datasets. Therefore, it is necessary to develop a strategy that addresses these challenges, which is reflected in the first sub-research question.

1. *How can reliable correspondences between road markings be established across LiDAR datasets with different acquisition times?*

As researched in Section 4.1, correspondences between overlapping LiDAR datasets with different acquisition times can be established by applying a stability-based method that filters road marking pairs using a RANSAC-based voting approach. The results show that this method generally manages to reduce misalignment, but in some tests, the method fails to do so, indicating a potential limitation of the proposed approach. In datasets with larger time gaps, the number of stable correspondences decreases, likely due to environmental changes. Moreover, the number of stable correspondences from automatically extracted road markings is consistently lower than from manually digitised ones, showing that extraction errors reduce the ability to establish reliable matches. Despite this condition, the automatically extracted road markings were still able to reduce the misalignment between overlapping datasets.

Based on the findings, the proposed method for establishing correspondences and aligning overlapping datasets appears to have certain limitations. To better understand these limitations, the second sub-research question was formulated.

2. *How does weighting the road markings based on their geometric stability influence the accuracy of multi-temporal LiDAR co-registration?*

As investigated in Section 4.2, the results show that weighting road markings based on their stability does not consistently improve alignment accuracy. Although the method can help exclude unstable correspondences and performs well when features are well-distributed, its advantage over a simpler nearest neighbour approach is minimal in most cases. The added complexity becomes a limitation when the number of initial correspondences is low or unevenly distributed, particularly in datasets with large time gaps. In these situations, the method tends to over-filter, reducing the influence of spatially important road markings and lowering alignment quality. Overall, while functional, the stability-based weighting does not always lead to better results.

As shown in the results from Section 4.1, there are factors in the automatically extracted road markings that affect the number of established correspondences, which also affect the alignment results when compared to manually digitised road markings. To better understand this impact, the third sub-research question was formulated.

3. *How does the accuracy of automatically extracted road markings influence the accuracy of LiDAR co-registration?*

The findings in Section 4.3 indicate that the accuracy of automatically extracted road markings influences co-registration results by affecting the distribution and reliability of the benchmarks used for alignment. Errors in the extraction process, such as undetected road markings and slight positional shifts of the road marking centroids, affect the alignment accuracy. However, the relationship is not straightforward. Errors in the extraction process affect the spatial distribution of the benchmarks, which then influences the alignment accuracy. Although the current method shows that extraction errors impact the alignment result, it does not measure how much they contribute to the overall error. While the current method shows that extraction errors have an effect on the alignment result, it does not quantify how large that effect is. This highlights a limitation of the proposed approach, as it cannot measure how much the extraction errors contribute to the overall alignment error.

The findings from the previous analysis also show that alignment errors vary across the test area. To better understand how these errors behave spatially and specifically how they change with increasing distance from the area where road markings are used as a benchmark, the fourth sub-research question was formulated.

4. *How does the alignment error change with increasing distance from the area where road markings are used as a benchmark?*

The findings show that alignment error increases with distance from the area where road markings are used as control features, especially when using automatically extracted road markings. This increase is mainly caused by rotation errors introduced during alignment, which become more noticeable further from the benchmark area. While manually digitised road markings produce stable alignment, automatically extracted features lead to larger misalignments over increasing distance. This suggests that differences in the transformation matrix produced by the alignment are influenced by the extraction quality and spatial distribution of the road markings. However, the current method does not establish a clear link between the amount of extraction error and the quality of distribution with the resulting rotation error. As a result, it is not possible to determine how specific levels of input error

lead to specific levels of transformation error, which limits the ability to fully understand the effect of these factors on alignment quality.

Besides the specific findings related to extraction accuracy and alignment results, this research also found uncertainties that can affect the position harmonisation process. These uncertainties do not happen in just one step but can appear at different points in the alignment process, from detecting the road markings to applying the transformation. To better highlight the uncertainties in the alignment process, the fifth sub-research question was formulated.

5. *What are the types of uncertainty that affect the accuracy of road marking extraction for position harmonisation?*

This research identified five main types of uncertainty that influence the accuracy of position harmonisation when using automatically extracted road markings. These uncertainties occur at different stages of the alignment process and include issues related to input data quality, extraction errors, correspondence estimation, transformation calculation, and accuracy validation. Each type contributes to reduced reliability in different ways, such as shifting road marking positions, reducing the number of valid correspondences, or distorting the final alignment result. A detailed summary of these uncertainties and their observed impacts is presented in Table 4.16, which highlights the challenges in using automatically extracted road markings for heterogeneous LiDAR alignment.

In response to the main research question, this research concludes that:

The findings of this research show that the suitability of automatically extracted road markings as benchmarks for LiDAR point cloud harmonisation is affected by both extraction errors and the spatial distribution of benchmarks. As the alignment time gap increases, environmental changes are more likely to occur between datasets, further reducing the number of available benchmarks. With fewer available benchmarks, the likelihood of an uneven benchmark distribution increases, which leads to errors in the estimated transformation. These transformation errors cause misalignments that become more noticeable with increasing distance from the benchmark area, resulting in higher alignment errors across the dataset.

Given these findings, it is not recommended to rely on automatically extracted road markings as benchmarks for co-registration in scenarios with large temporal gaps between datasets, especially in cases where the road markings are only available from a single highway that runs in a single axis. In such cases, the risk of having both a low number of available control features and an uneven spatial layout is significantly higher, making the resulting transformation less reliable. It is also not recommended to use areas located far from the benchmark region in further processing or spatial analysis, especially in cases with large alignment time gaps, as alignment errors tend to increase with distance from the control area. These conclusions are only applicable to cases similar to the study presented in this research, as different conditions may lead to different outcomes.

However, a limitation of the proposed method is that it does not quantify how specific levels of extraction error or distribution quality propagate into transformation errors. This limits the ability to fully understand or predict the effects of those input conditions on the final alignment quality. Additionally, this research only investigates one type of road layout, where all road markings run in a single direction (along the x-axis). As a result, the impact of other spatial configurations, such as intersections or curved roads, could not be evaluated, which may influence the generalisability of the findings.

Reflection of the research:

The findings of this research provide insight into the performance of the current development of automatic road marking extraction as a benchmark for heterogeneous LiDAR point cloud harmonisation, as introduced in Heide 2024, specifically within the study case proposed in this research. This benchmark has the potential to support the goals of the Integrale Hoogtevoorziening Nederland (IHN) initiative, which aims to co-register point cloud data across the Netherlands.

The research also proposes a scheme to observe the relationship between the quality of extracted benchmarks and the resulting alignment accuracy. This scheme provides insight into which aspects of the relationship can be effectively observed, while also revealing its own limitations, specifically where it fails to capture deeper patterns needed for more influential conclusions, such as its inability to quantify how much a specific input condition contributes to the resulting alignment error. These observations help identify directions for improving the evaluation method itself. In the context of geomatics, this is important in improving the quality assessment of point cloud data, as accurate point cloud alignment plays a key role in supporting reliable 3D modelling and spatial analysis.

In addition, the research proposes a new strategy for establishing correspondences and estimating transformations based on the geometric stability of individual features. This strategy was also compared with a simpler, more direct method to evaluate its effectiveness. The comparison not only informs the performance of the proposed strategy but also contributes new insight into feature-based alignment strategies for heterogeneous point cloud harmonisation.

The code used to compute the geometry feature thresholds for the road marking cluster estimation step, to establish correspondences using the RANSAC inlier voting method, and to perform alignment using the RANSAC-weighted centroid method, as well as the training datasets used to compute the geometry feature thresholds and the sample datasets used for alignment in this research, are available in the following GitHub repository: Research Repository.

6.2 Recommendations

The following recommendations are proposed to address the limitations and challenges identified in this study and to support future improvements in similar research.

- **Improved filtering of non-road marking objects:** The automatic extraction often included false features (e.g., road edge lines), which negatively affected the alignment. Enhanced filtering techniques during extraction and in the correspondence establishment phase could help remove these false positives.
- **Improved extracted line accuracy:** In this study, some of the extracted road markings did not match the true position or shape of the actual markings. This inaccuracy indirectly reduced the quality of the alignment. Future work can investigate improving how the lines are extracted, such as by improving the line segmentation method.
- **Investigation of influence of benchmark distribution on transformation matrix values:** Further investigation is needed to understand the influence of spatial benchmark distribution on specific transformation components (e.g., omega, phi, kappa). This

could help to determine the extent to which benchmark distributions can be classified as stable or unstable for point cloud alignment purposes. Controlled experiments could be carried out to simulate different benchmark distribution scenarios.

- **Dynamic weighting of benchmark:** Instead of relying on uniformly weighted benchmarks, future work could explore adaptive weighting strategies that consider the initial spatial distribution of the benchmarks as input to give weight to the benchmarks.
- **Combining different types of benchmarks:** Combining road markings with other types of benchmarks (building surfaces, railway features, lamp posts, etc.) may help increase the number of benchmarks, which reduces the risk of unstable distribution of benchmarks
- **Incorporating RGB values for improved road markings filtering:** Some high-intensity non-road marking features, such as unfiltered ground-level car points or planted steel on roads, may share similar intensity values with road markings and lead to misclassification. Integrating RGB values can provide additional information to help distinguish road markings from other high-intensity objects. This could improve both the filtering of false positives of the extracted benchmark.

Bibliography

Ackroyd, Chelsea et al. (2024). "Airborne lidar intensity correction for mapping snow cover extent and effective grain size in mountainous terrain". In: *GIScience & Remote Sensing* 61.1, p. 2427326. DOI: 10.1080/15481603.2024.2427326. eprint: https://doi.org/10.1080/15481603.2024.2427326. URL: https://doi.org/10.1080/15481603.2024.2427326.

Actueel Hoogtebestand Nederland (AHN) Congress (2024). *Congres AHN & Beeldmateriaal 2024*. Accessed: 6 March 2025. URL: https://www.ahn.nl/congres-ahn-beeldmateriaal-2024.

AHN (2024). *About the AHN programme*. Accessed: 2025-04-08. URL: https://www.ahn.nl/en/about-ahn.

AHN Dataroom (2024). *Data room – AHN*. URL: https://www.ahn.nl/en/data-room.

Arun, K. S., T. S. Huang, and S. D. Blostein (1987). "Least-Squares Fitting of Two 3-D Point Sets". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9.5, pp. 698–700. DOI: 10.1109/TPAMI.1987.4767965.

Barçon, E. et al. (2022). "EXTRACTION OF ROAD MARKINGS FROM MLS DATA: A REVIEW". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-2/W1-2022, pp. 7–14. DOI: 10.5194/isprs-archives-XLVIII-2-W1-2022-7-2022. URL: https://isprs-archives.copernicus.org/articles/XLVIII-2-W1-2022/7/2022/.

Benallal, Hafsa et al. (2022). "A new approach for removing point cloud outliers using the standard score". In: *Pattern Recognition and Tracking XXXIII*. Ed. by Mohammad S. Alam and Vijayan K. Asari. Vol. 12101. International Society for Optics and Photonics. SPIE, p. 1210107. DOI: 10.1117/12.2618835. URL: https://doi.org/10.1117/12.2618835.

Besl, P.J. and Neil D. McKay (1992). "A method for registration of 3-D shapes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2, pp. 239–256. DOI: 10.1109/34.121791.

Brightman, Nathan, Lei Fan, and Yang Zhao (2023). "Point cloud registration: a mini-review of current state, challenging issues and future directions". In: *AIMS Geosciences* 9.1, pp. 68–85. ISSN: 2471-2132. DOI: 10.3934/geosci.2023005. URL: https://www.aimspress.com/article/doi/10.3934/geosci.2023005.

Cheng, Liang et al. (2018). "Registration of Laser Scanning Point Clouds: A Review". In: *Sensors* 18.5. ISSN: 1424-8220. DOI: 10.3390/s18051641. URL: https://www.mdpi.com/1424-8220/18/5/1641.

Cheng, Yi-Ting et al. (2020). "Intensity Thresholding and Deep Learning Based Lane Marking Extraction and Lane Width Estimation from Mobile Light Detection and Ranging (LiDAR) Point Clouds". In: *Remote Sensing* 12.9. ISSN: 2072-4292. DOI: 10.3390/rs12091379. URL: https://www.mdpi.com/2072-4292/12/9/1379.

Cheng, Xiaolong et al. (2018). "Automatic Registration of Terrestrial and Airborne Point Clouds Using Building Outline Features". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.2, pp. 628–638. DOI: 10.1109/JSTARS.2017.2788054.

Csanyi, Nora and Charles K Toth (2007). "Improvement of lidar data accuracy using lidar-specific ground targets". In: *Photogrammetric Engineering & Remote Sensing* 73.4, pp. 385–396.

Dekking, F.M. et al. (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer. URL: https://link.springer.com/book/10.1007/1-84628-168-7.

Donkers, T.A.W. (July 2024). "Assessing the Quality of LiDAR Infrastructure Point Clouds". Available at: http://repository.tudelft.nl/. MA thesis. Delft, The Netherlands: Delft University of Technology.

Fischler, Martin A. and Robert C. Bolles (1987). "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". In: *Readings in Computer Vision*. Ed. by Martin A. Fischler and Oscar Firschein. San Francisco (CA): Morgan Kaufmann, pp. 726–740. ISBN: 978-0-08-051581-6. DOI: https://doi.org/10.1016/B978-0-08-051581-6.50070-2. URL: https://www.sciencedirect.com/science/article/pii/B9780080515816500702.

Fontana, Simone et al. (June 2021). "A benchmark for point clouds registration algorithms". In: *Robotics and Autonomous Systems* 140, p. 103734. ISSN: 0921-8890. DOI: 10.1016/j.robot.2021.103734. URL: http://dx.doi.org/10.1016/j.robot.2021.103734.

Gao, Yang et al. (July 2017). "Automatic extraction of pavement markings on streets from point cloud data of mobile LiDAR". In: *Measurement Science and Technology* 28.8, p. 085203. DOI: 10.1088/1361-6501/aa76a3. URL: https://dx.doi.org/10.1088/1361-6501/aa76a3.

GeoTiles.nl (2024). *GeoTiles — Ready-made geodata with a focus on the Netherlands*. URL: https://www.geotiles.nl/.

Gillies, Sean and contributors (2013). *Rasterio: access to geospatial raster data*. https://github.com/rasterio/rasterio. Version 1.3+, accessed 2025-05-13.

Habib, AF et al. (2008). "LiDAR strip adjustment using conjugate linear features in overlapping strips". In: *Proceedings of International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 385–390.

Harris, Charles R. et al. (2020). "Array Programming with NumPy". In: *CoRR* abs/2006.10256. arXiv: 2006.10256. URL: https://arxiv.org/abs/2006.10256.

Heide, Daan van der (Dec. 2024). "The harmonisation of heterogeneous point clouds into one national dataset: Towards an Integral Height-dataset of the Netherlands (IHN)". PhD Candidate 2023–2028. PhD Proposal. TU Delft.

Heide, Daan van der et al. (Aug. 2024). *WP1: Inventarisatie van puntenwolken in Nederland*. Tech. rep. Versie 1.0. Accessed 2025-03-25. Rijkswaterstaat, Het Waterschapshuis, TU Delft.

Heipke, Christian et al. (1997). "Evaluation of automatic road extraction". In: *International Archives of Photogrammetry and Remote Sensing* 32.3 SECT 4W2, pp. 151–160.

Horn, Berthold K. P. (Apr. 1987). "Closed-form solution of absolute orientation using unit quaternions". In: *J. Opt. Soc. Am. A* 4.4, pp. 629–642. DOI: 10.1364/JOSAA.4.000629. URL: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-4-4-629.

Horn, Grant Van, Joseph C. Hruska, and contributors (2023). *Laspy: A Python library for reading, modifying, and writing LAS LiDAR files*. https://github.com/laspy/laspy. Version 2.4.1, Accessed: 2025-05-13.

Huang, Xiaoshui et al. (2021). *A comprehensive survey on point cloud registration*. arXiv: 2103.02690 [cs.CV]. URL: https://arxiv.org/abs/2103.02690.

Kalenjuk, Slaven and Werner Lienhart (2022). "A Method for Efficient Quality Control and Enhancement of Mobile Laser Scanning Data". In: *Remote Sensing* 14.4. ISSN: 2072-4292. DOI: 10.3390/rs14040857. URL: https://www.mdpi.com/2072-4292/14/4/857.

Keuris, Lars (Nov. 2024). *Nationaal Noklijnenbestand*. Tech. rep. Version 1.0. Commissioned by Het Waterschapshuis as part of the Integrale Hoogtevoorziening Nederland (IHN). Netherlands: Keuris Earth Observation. URL: https://earth.observation.keuris.nl.

Killinger, D.K. (2014). "10 - Lidar (light detection and ranging)". In: *Laser Spectroscopy for Sensing*. Ed. by Matthieu Baudelet. Woodhead Publishing, pp. 292–312. ISBN: 978-0-85709-273-1. DOI: https://doi.org/10.1533/9780857098733.2.292. URL: https://www.sciencedirect.com/science/article/pii/B9780857092731500106.

Kim, Minsu et al. (2022). "Absolute Accuracy Assessment of Lidar Point Cloud Using Amorphous Objects". In: *Remote Sensing* 14.19. ISSN: 2072-4292. DOI: 10.3390/rs14194767. URL: https://www.mdpi.com/2072-4292/14/19/4767.

Kushwaha, Sunni Kanta Prasad et al. (2023). "Integrating Airborne and Terrestrial Laser Scanning for Complete 3D Model Generation in Dense Forest". In: *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 3137–3140. DOI: 10.1109/IGARSS52108.2023.10283032.

Leusink, Jeroen (Oct. 2024). *Integrale Hoogtevoorziening Nederland (IHN)*. Presentation at Congres AHN & Beeldmateriaal 2024, De Mariënhof, Amersfoort. URL: https://www.ahn.nl/congres-ahn-beeldmateriaal-2024.

Li, Xiaolu et al. (2023). "LiDAR intensity correction for road marking detection". In: *Optics and Lasers in Engineering* 160, p. 107240. ISSN: 0143-8166. DOI: https://doi.org/10.1016/j.optlaseng.2022.107240. URL: https://www.sciencedirect.com/science/article/pii/S0143816622002937.

Liang, Yu-Bin et al. (2014). "Automatic Registration of Terrestrial Laser Scanning Data Using Precisely Located Artificial Planar Targets". In: *IEEE Geoscience and Remote Sensing Letters* 11.1, pp. 69–73. DOI: 10.1109/LGRS.2013.2246134.

Mi, Xiaoxin et al. (2021). "A two-stage approach for road marking extraction and modeling using MLS point clouds". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 180, pp. 255–268. ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2021.07.012. URL: https://www.sciencedirect.com/science/article/pii/S0924271621001970.

ProRail (2024). *ProRail*. Accessed: 2025-04-08. URL: https://www.prorail.nl/over-ons/wat-doet-prorail.

ProRail SpoorInBeeld (2024). *SpoorInBeeld – SpoorData*. URL: https://www.spoordata.nl/informatieproducten/spoorinbeeld.

Rijkswaterstaat (n.d.). *About us - Rijkswaterstaat*. Accessed: 2025-04-08. URL: https://www.rijkswaterstaat.nl/en/about-us.

Riofrío, José et al. (2022). "Harmonizing multi-temporal airborne laser scanning point clouds to derive periodic annual height increments in temperate mixedwood forests". In: *Canadian Journal of Forest Research* 52.10, pp. 1334–1352. DOI: 10.1139/cjfr-2022-0055. eprint: https://doi.org/10.1139/cjfr-2022-0055. URL: https://doi.org/10.1139/cjfr-2022-0055.

Rodríguez-Gonzálvez, Pablo et al. (2017). "Mobile LiDAR System: New Possibilities for the Documentation and Dissemination of Large Cultural Heritage Sites". In: *Remote Sensing* 9.3. ISSN: 2072-4292. DOI: 10.3390/rs9030189. URL: https://www.mdpi.com/2072-4292/9/3/189.

Salach, Adam et al. (2018). "Accuracy Assessment of Point Clouds from LiDAR and Dense Image Matching Acquired Using the UAV Platform for DTM Creation". In: *ISPRS International Journal of Geo-Information* 7.9. ISSN: 2220-9964. URL: https://www.mdpi.com/2220-9964/7/9/342.

Schenk, T (2001). "Modeling and analyzing systematic errors in airborne laser scanners". In: *Technical notes in Photogrammetry* 19.46.

Shahraji, M. H., C. Larouche, and M. Cocard (2020). "ANALYSIS OF SYSTEMATIC ERRORS OF MOBILE LiDAR SYSTEMS: A SIMULATION APPROACH". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-1-2020, pp. 253–260. DOI: 10.5194/isprs-annals-V-1-2020-253-2020. URL: https://isprs-annals.copernicus.org/articles/V-1-2020/253/2020/.

Shao, Jie et al. (2022). "Efficient co-registration of UAV and ground LiDAR forest point clouds based on canopy shapes". In: *International Journal of Applied Earth Observation and Geoinformation* 114, p. 103067. ISSN: 1569-8432. DOI: https://doi.org/10.1016/j.jag.2022.103067. URL: https://www.sciencedirect.com/science/article/pii/S1569843222002552.

Soudarissanane, S (2016). "The geometry of terrestrial laser scanning; identification of errors, modeling and mitigation of scanning geometry". In: *TU Delft*.

Tiberius, CC et al. (2021). "Surveying and mapping. TU Delft Open". In: *Delft University of Technology, The Netherlands*.

Umeyama, S. (1991). "Least-squares estimation of transformation parameters between two point patterns". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13.4, pp. 376–380. DOI: 10.1109/34.88573.

Urbančič, Tilen et al. (2019). "New Target for Accurate Terrestrial Laser Scanning and Unmanned Aerial Vehicle Point Cloud Registration". In: *Sensors* 19.14. ISSN: 1424-8220. DOI: 10.3390/s19143179. URL: https://www.mdpi.com/1424-8220/19/14/3179.

Vosselman, George (2008). "Analysis of planimetric accuracy of airborne laser scanning surveys". In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37.3a, pp. 99–104.

Wang, Yanjun et al. (2019). "A Survey of Mobile Laser Scanning Applications and Key Techniques over Urban Areas". In: *Remote Sensing* 11.13. ISSN: 2072-4292. DOI: 10.3390/rs11131540. URL: https://www.mdpi.com/2072-4292/11/13/1540.

Wang, Yanmin et al. (2014). "Automatic registration of laser point cloud using precisely located sphere targets". In: *Journal of Applied Remote Sensing* 8.1, p. 083588. DOI: 10.1117/1.JRS.8.083588. URL: https://doi.org/10.1117/1.JRS.8.083588.

Wen, Chenglu et al. (2019). "A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 147, pp. 178–192. ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2018.10.007. URL: https://www.sciencedirect.com/science/article/pii/S0924271618302855.

Wu, Hangbin and Hongchao Fan (2016). "Registration of Airborne LiDAR Point Clouds by Matching the Linear Plane Features of Building Roof Facets". In: *Remote Sensing* 8.6. ISSN: 2072-4292. DOI: 10.3390/rs8060447. URL: https://www.mdpi.com/2072-4292/8/6/447.

Xiao, WEN (2012). "Detecting changes in trees using multi-temporal airborne LIDAR point clouds". MA thesis. University of Twente.

Xiaohong, Zhang and Liu Jingnan (Jan. 2004). "Analysis of systematic error influences on accuracy of airborne laser scanning altimetry". In: *Geo-spatial Information Science* 7.3, pp. 218–224. DOI: 10.1007/bf02826295.

Xu, Wenxue et al. (2022). "Feature curve-based registration for airborne LiDAR bathymetry point clouds". In: *International Journal of Applied Earth Observation and Geoinformation* 112, p. 102883. ISSN: 1569-8432. DOI: https://doi.org/10.1016/j.jag.2022.102883. URL: https://www.sciencedirect.com/science/article/pii/S1569843222000851.

Yan, Li et al. (2016). "Scan Line Based Road Marking Extraction from Mobile LiDAR Point Clouds". In: *Sensors* 16.6. ISSN: 1424-8220. DOI: 10.3390/s16060903. URL: https://www.mdpi.com/1424-8220/16/6/903.

Yang, Bisheng et al. (2012). "Automated extraction of road markings from mobile LiDAR point clouds". In: *Photogrammetric Engineering & Remote Sensing* 78.4, pp. 331–338.

Yang, Ronghao et al. (2020). "Accurate Road Marking Detection from Noisy Point Clouds Acquired by Low-Cost Mobile LiDAR Systems". In: *ISPRS International Journal of Geo-Information* 9.10. ISSN: 2220-9964. DOI: 10.3390/ijgi9100608. URL: https://www.mdpi.com/2220-9964/9/10/608.

Yao, Lianbi et al. (2021). "Automatic Road Marking Extraction and Vectorization from Vehicle-Borne Laser Scanning Data". In: *Remote Sensing* 13.13. ISSN: 2072-4292. DOI: 10.3390/rs13132612. URL: https://www.mdpi.com/2072-4292/13/13/2612.

Zhang, Wuming et al. (2016). "An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation". In: *Remote Sensing* 8.6. ISSN: 2072-4292. DOI: 10.3390/rs8060501. URL: https://www.mdpi.com/2072-4292/8/6/501.

Colophon

This document was typeset using LaTeX, using the KOMA-Script class `scrbook`. The main font is Palatino.