UNIVERSITÀ DEGLI STUDI DI PADOVA

Corso di laurea in Scienze Forestali e Ambientali oppure

Corso di laurea in Tecnologie Forestali e Ambientali

Tree detection and geolocalization with depth-camera and deep learning algorithms

Supervisor:

Luca Marchi, Professor, University of Padova

Co-Supervisors:

Stefano Grigolato, Professor, University of Padova

Christopher Wong, Professor, University of New Brunswick

Gunta Grube, PhD Candidate, University of Padova

<div align="right">

Graduating Student:

Alicia Sood

Enrollment Nº: 2105063

July 8, 2024

</div>

Academic Year: 2023-2024

Summary

A new approach of low-cost terrestrial mapping sensors is the ZED 2i stereo camera, which has been introduced into forestry settings. A stereo camera simulates human vision, capturing two simultaneous images that can be used to estimate depth and motion of detected objects. This includes training the necessary detection software to run with the ZED 2i Stereo Camera, collecting data in varying environments, converting the local coordinates of detected trees to a global coordinate system, identifying the more reliable observations, and consolidating clusters of observations for same trees into a singular tree position. Studies were conducted on the overall performance of the stereo camera compared to extracted control data from drone orthophotos, and saw that more observations were detected in a forestry plantation when foliage had developed in May versus February tests when there were no leaves on the trees. Testing also compared data results from a tree plantation to a more complex, natural forest environment, which saw a reduction in overall observation retention after being put through accuracy filtering based on detection confidence and distance of detections from the camera.

In brief, data was used to configure processing procedures to allow for the extraction of meaningful results (consolidated tree positions, comparison of different test path types and different test environments), which allowed for an opening assessment of the performance of the processing methods established against control data gathered in the same environments using established methods such as GNSS positioning or drone data extraction. Forestry processes are calling for the implementation of new technologies for automatic detection, and automated vehicles and machinery. This thesis lays the groundwork for future improvements in data interpretation and accuracy based on the developed procedures, highlighted assumptions, limitations, and next steps for working to implement the use of a ZED 2i Stereo Camera in the forestry industry.

List of Contents

List of Figures

List of Tables

List of Equations

1.0 Introduction

In forested areas, terrestrial mapping methods can be much more effective than aerial photogrammetry (McGlade et al., 2022). There are different approaches, including terrestrial "manual" land surveys using a global navigation satellite system (GNSS), or highly accurate point clouds using laser scanners and Light Detection and Ranging (LiDAR) systems.

When using GNSS, a consideration when it comes to positioning is the obtainable accuracy, which can greatly depend on the environment (e.g., if it is being obstructed by a thicker canopy, etc.) (Wells & Chung, 2023). Terrestrial laser scanning (TLS) can attain a higher degree of accuracy but take a longer time to scan the area (Lagisetty et al., 2013) as it operates for a period of time from fixed points. LiDAR, rather than collecting precision data from a stationary position, is attached to a moving platform such as a drone or vehicle (McGlade et al., 2023). These remote sensing techniques have become extremely valuable in the forestry setting, increasingly replacing more manual surveying methods such as measurements with calipers, etc. (McGlade et al., 2023). While optimal in theory due to rapid and accurate measurements of environmental characteristics, laser scanners and LiDAR both come at a greater cost for their precision capabilities (Kao et al., 2022) (McGlade et al., 2023). Accurate mapping is important for endless purposes; however, the high base cost of equipment can be prohibitive for many smaller companies and organizations, creating a barrier to entry in industries dominated by larger corporations and government bodies.

There are a variety of new approaches for more cost-effective three-dimensional remote sensing mapping which typically include a red, green, and blue colour depth (RGB-D) sensors and an inertial measurement unit (IMU) (McGlade et al., 2023). Available applications of such technologies include identifying the existence of objects/obstacles (Arunpriyan et al., 2020) and evaluating position and dimensions (Nirunsin et al., 2022), so specific equipment can be chosen based on research or work needs as there are a variety of options on the market. For example, McGlade et al. (2023) explored the accuracy of both the Stereolab Zed 2 camera and the Apple iPad Pro 2020, highlighting their strengths and weaknesses depending on the intended application.

Stereo vision incorporates multiple lenses, simulating human vision in its image and depth capturing (Liszka et al., 2023). It offers that lower cost mapping option, meeting the current interest and demand for affordable and broadly applicable mapping equipment (Lagisetty et al., 2013).

With its camera lenses typically placed at the average human eye width apart (6.5cm), the camera can capture two images at the same time that are slightly offset (Fig. 1.1). The pixel displacement between the left and right images can then be used to estimate depth and motion (Stereolabs Inc., 2024b).
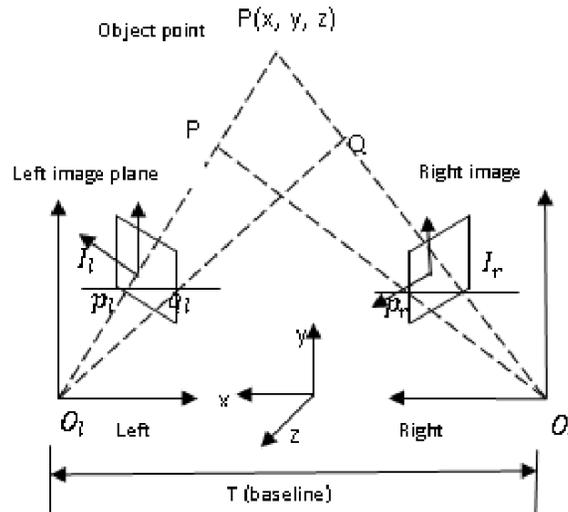
Figure 1.1 The Geometry of a stereo camera geometry of a stereo camera (Lagisetty et al., 2013)

Image rectification takes the two images from the stereo camera, and converts them into one common reference frame, while triangulation then compares the images, using the baseline between lenses and their focal length in order to triangulate the 3d coordinates of a point (Lagisetty et al., 2013)(Liszka et al., 2023).

Many industries are exploring this ZED stereo camera technology for a multitude of implementations, for example Almendral et al., (2018) implemented its system with respect to fruit harvesting. It was used for depth estimation of automated arms to be used for picking or cutting crops. They concluded that the stereo camera was effective for capturing longer ranges of up to 20m, while it was found not so effective in ranges closer than 65 cm to the camera. Juang et al. (2016) implemented use of a stereo camera for detection of objects containing multiple colors with non-homogeneous distributions in complex backgrounds to estimate objects' depth and shape.

At present, there is also extensive ongoing research and development on autonomous road vehicles. (Li & Lideskog, n.d.). Lagisetty et al. (2013) used stereo cameras to research obstacle avoidance and object detection for a mobile robot using a stereo camera geared towards

researching collision avoidance in autonomous robotic systems. Arunpriyan et al. (2020) explicitly highlights the improvement a stereo camera offers from a monocular camera stating that using a stereo vision camera with depth information instead of a monocular camera, the overall accuracy of real-time speed bump detection using image segmentation from autonomous vehicles was improved. Despite this increase in research for autonomous vehicles, not as much focus is put explicitly on autonomous driving in off-road environments (Li & Lideskog, n.d.).

Looking specifically at the forestry industry, many forestry processes affect ecosystem health and work efficiency, calling for the implementation of new technologies (Liszka et al., 2023). This has significantly contributed to the drive for more autonomous vehicles and machinery, as well as automatic detection. This can enhance the production efficiency of forestry operations by using detection for navigational purposes and facilitating removal processes (Li & Lideskog, n.d.). For example, Li & Lideskog aimed to improve productivity of work by implementing automatic detection of stumps and boulders to better avoid or tackle obstructions to improve work efficiency. Liszka et al. (2023) looked to evaluate whether trees were successfully planted by autonomous vehicles, aiming to improve research towards potential sustainable land management practice implementation. Fleischmann & Berns (n.d.) pursued the effectiveness of stereo cameras for agricultural applications, making several cross-discipline conclusions regarding stereo camera technology. This includes its abilities to provide a dense point cloud of information as well as colour information, and its convenience of light-weight nature, low cost, and low energy consumption. In testing it was also concluded that influence of dust with stereo camera performance was lower than for a laser scanner (Fleischmann & Berns, n.d.), therefore this may indicate its potential benefits in applying in forest and harvesting settings with the presence of saw dust, pollen, etc. This is a topic of particular interest for the long-term goals that this thesis will build foundation for, the implementation of stereo camera vision for object detection and obstacle avoidance in forest operations.

Given this basis of interest in expanding the use of low-cost stereo vision, it is of interest to acquire and investigate the process of setting up, using, and processing data using this technology to open doors for abundant potential research and work opportunities.

1.1 Objectives

There are numerous studies on using stereo cameras together with advanced algorithms, to identify and gather dimensional data about objects in its field of view, including size, shape and position of the object. This work will train ZED 2i Stereo Camera detection software with acceptable accuracy for object detection and sizing in forested areas in Italy in order to configure processing procedures to allow for the extraction of meaningful insights from the amassed data.

The specific objectives to complete this general objective are as follows:

o Achieve functioning, trained detection software to implement with the ZED 2i Stereo Camera for data collection of tree positions.

o Achieve conclusive surveys of a plantation test area (Legnaro, PD) without and with tree foliage; tree detections positioned in local coordinate systems with respect to the camera position.

o Achieve conclusive surveys of a natural forest test area (Taibon Agordino, BL); tree detections positioned in local coordinate systems with respect to the camera position.

o Construct methodology for processing procedures of stereo camera detections and Global Positioning System (GPS) position data.

o Analyse the performance of the processing methods established against control data (from GPS or drone data).

o Discuss the extent to which environmental considerations impact accuracy and quantity of detection and to conclude whether any error incurred by environmental considerations is significant.

2.0 Materials

This section provides a detailed breakdown of all resources required for each phase of the thesis work, including preliminary planning, data collection, processing, and finalizing results. This encompasses different software to run the necessary programs, equipment needed to conduct the field surveys, and any additional materials.

2.1 Software

2.1.1 ZED SDK

A software development kit (SDK) is a compilation of necessary software tools compiled into a library. The composed ZED SDK will facilitate the usage of the ZED stereo camera with compatible languages, platforms, and third-party integrations. This will include YOLO, Python, OpenCV, and PyTorch.

2.1.2 YOLO

You Only Look Once (YOLO) is a machine learning algorithm for object detection that will take the stereo images that provide depth information and identify objects within the photo. With many models and application techniques available, da Silva et al. (2022) provided both a comparison of varying YOLO models, as well as evidence for use with specific regard to tree mapping. Images are gathered and used to train the model to identify key features of interest (Jrondi et al., 2024). While there are a variety of machine learning algorithms to pair with stereo camera work, YOLOv8 provides a tested and reliable performance (Liszka et al., 2023). This thesis work specifically uses YOLOv8 developed by Ultralytics, LLC as a compatible third-party library with ZED SDK.

2.1.3 Python

Python is a programming language maintained by the Python Software Foundation (PSF). This thesis work specifically uses Python version 3.8 for writing and running the code required by the stereo camera and simultaneous usage and recording of independent GPS data.

2.1.4 OpenCV

Open Source Computer Vision Library (OpenCV) is an open-source computer vision and machine learning software library (OpenCV, 2024). This thesis work specifically uses OpenCV

version 4.9.0 for image processing and support of the other deep learning frameworks in use as a compatible third-party library with ZED SDK.

2.1.5 PyTorch

PyTorch is also an open-source computer vision and machine learning library (PyTorch, 2024). This thesis work specifically uses PyTorch version 1.8.0 with CUDA 1.1.1 for image processing and support of the other deep learning frameworks in use as a compatible third-party library with ZED SDK.

2.1.6 R

R is a software for statistical computing and graphics (R-Project, 2024). This thesis work specifically uses R version 4.4.0 and RStudio 2024.04.2 for data computing, statistical analysis, and visualizations.

2.1.7 Excel

Microsoft Excel is a spreadsheet software developed and maintained by Microsoft Corporation. This thesis work specifically uses Excel for the organization, amalgamation, and calculations of post processing steps for tree observation Data.

2.1.8 QGIS

QGIS is an open-source geographic information system and will be the primary processing and mapping tool for the geospatial information of this thesis (*QGIS*, 2024). Version 3.28.12 will be used for the visualization and comparison of results, enabling quantitative evaluations of the relationship between datasets.

2.2 Equipment

2.2.1 ZED 2i Stereo Camera

For Field Surveys 1 and 3, the stereo camera implemented is the ZED 2i Stereo Camera, with a 2.1mm fixed focal lens and polarizing filter to improve the quality of outdoor testing. For Field Survey 2, the stereo camera acquired is the ZED 2i Stereo Camera, with a 4mm fixed focal lens and polarizing filter to improve the quality of outdoor testing. The difference (due to availability of equipment in the department) is the 2.1mm lens provides an ultra wide field of view

with optically corrected distortion for increase image quality, while the 4mm lens provides increased resolution and depth accuracy at longer range (Fig. 2.1) (STEREOLABS, 2022).

Figure 2.1 Stereo camera difference between 2.1mm fixed focal lens and 4mm fixed focal lens (STEREOLABS, 2022)

These two variations of the ZED 2i Stereo Camera hold distinct specifications (Table 2.1). Camera output resolutions and output files are presented in Table 2.2. and 2.3 respectively.

Table 2.1 Specifications of the ZED 2i Stereo Camera with a 2.1mm and 4mm fixed focal lens

Lens length	Sensor Type	Frame rate	Array Size	Focal Length	Field of View	Aperture	Depth Range	Depth Accuracy	Object Detection
2.1mm				2.12mm (0.008")	H: 110° V: 70° D: 120°	f/1.8	0.3 - 20 m (1 - 65.6ft)	< 1% up to 3m < 5% up to 15m	Up to 20m (3D) Up to 40m (2D)
4mm	1/3" 4MP CMOS	15-100 FPS	2688 x 1520 pixels	4mm (0.16")	H: 72° V: 44° D: 81°	f/1.8	1.5 - 35 m (4.9 - 115ft)	< 2% up to 10m < 7% up to 30m	Up to 35m (3D) Up to 55m (2D)

Table 2.2 Output Resolution of a ZED 2i Stereo Camera

Resolution	Frame Rate	Mode
2x (2208x1242)	15 FPS	cropping mode
2x(1920x1080)	15/30 FPS	cropping mode
2x (1280x720)	15/30/60 FPS	binning 2x2 mode
2x(662x376)	15/30/60/100 FPS	binning 4x4 mode

Table 2.3 Output Files and file components of a ZED 2i Stereo Camera (Stereolabs Inc., 2024a)

Files	Information
All Detections	**Object ID**: An identifier for each detected object. **Label**: Classification of the object based on trained object types. **Position**: The 3D position of an identified object relative to the camera. **Velocity**: 3D vector of an object's movement according to the camera. **Dimensions**: The length, width, and height of the object. **Confidence**: The model's confidence level in the object's detection and classification (0-100). **Tracking State**: Whether the object is tracking or lost (Ok, Off, Searching, Terminate). **Action State**: Whether the object is moving or idle. **Distance to Cam:** The distance from the camera to the object. **Detection Timestamp**: The time at which the object was detected. **Roll:** The camera's rotation around the front-to-back (Z) axis. **Pitch**: The camera's rotation around the side-to-side (X) axis. **Yaw**: The camera's rotation around the vertical (Y) axis.
Closest Detections	Same as "All Detections" information, reduced to one singular object observations per group of observations assigned the same object ID when the camera was closets to the observation.
Output Video	- Video recording of the observation period - Simultaneous 3D object tracking of the identified objects in the environment, displaying objects' path over time.
Recorded GPS Data	**Timestamp**: The time at which the camera position is recorded. **Latitude**: The geographic coordinates indicating the camera's position. **Longitude**: The geographic coordinates indicating the camera's position. **Altitude**: The elevation of the camera above sea level. **Satellites**: Number of connected satellites. **HDOP**: Horizontal Dilution of Precision, geometric strength of satellite configuration. **Geoid Height**: The vertical distance between the geoid and the reference ellipsoid.

A sample of the video output from the stereo camera, which can also be observed live during testing, is presented in Figure 2.2.

Figure 2.2 Sample video output screenshot from the ZED 2i Stereo Camera, showing real-time object identification (left) and 3D object tracking (right)

2.2.2 GNSS (Emlid Reach RS2 and Emlid Reach 2)

A GNSS real-time kinematic (RTK) approach was used in conjunction with the stereo camera to improve object positioning. This used a network of a reference station and a rover receiver to provide highly accurate and precise positioning in real-time, providing centimeter-level accuracy (Stereolabs Inc., 2024d). Obstructions and atmospheric errors hinder GPS accuracy, therefore implementing a base station nearby allowed for the real-time correction of errors in satellite signals for a rover paired with the stereo camera for stereo camera positioning, with shorter baselines allowing for increased accuracy (Na'aim & Manaf, 2024).

An Emlid Reach RS2 receiver was used as a base station during each survey to enhance the accuracy of positioning (Emlid, 2022). An Emlid Reach M2 was used as the rover for setting control in the test environments, as well as collecting GPS data as a reference survey for later comparison with stereo camera data and mounted on the stereo camera to provide coordinates of the camera's zero position when assigning local coordinates to observed objects during the tests (Emlid, 2019). Fixed positional accuracies of these GNSS systems are summarized in Table 2.4, while float can reduce this centimetre-level accuracy to decimeter-level precision (OXTS, 2020).

Table 2.4 Positioning accuracy of the Emlid Reacg RS2 and Emlid Reach M2

	Emlid Reach RS2	Emlid Reach M2
Static	H: 4 mm + 0.5 ppm V: 8 mm + 1 ppm	H: 4mm + 0.5 ppm V: 8 mm + 1 ppm
RTK	H: 7 mm + 1 ppm V: 14 mm + 1 ppm	H: 7 mm + 1 ppm V: 14 mm + 1 ppm

These units record raw positioning data in RINEX format for subsequent post-processing. Data logged by the rover will include elements summarized in Table 2.5.

Table 2.5 Output information logged from Emlid Reach M2

Files	Information
RINEX Output File	**Date**: The date on which the position was recorded. **Timestamps**: The time at which the position was recorded. **Latitude**: The geographic latitude of the recorded position. **Longitude:** The geographic longitude of the recorded position. **Ellipsoidal Height:** The height above the reference ellipsoid **Solution:** The type of GNSS solution (e.g., fixed, float, single) **# Satellites:** The number of satellites used in the solution. **Field 8-15:** Positional accuracy metrics.

2.2.3   Laptops

A laptop was used for training detection software and the actual software use in conjunction with the stereo camera during the tests. During testing, it is connected to both the ZED 2i stereo camera and rover, as well as running the YOLOv8 model for object detection, depth sensing, and all recording of data.

The ZED SDK software development kit designed for use with ZED stereoscopic cameras can operate on any Windows or Linux platforms, however it has recommended minimum specifications (Fig. 2.3) which must be considered in the selection of a laptop for training and running a stereo camera. Additionally, a NVIDIA® GPU with Compute Capabilities greater than 3 is required (Stereolabs Inc., 2024c).

	MINIMUM	RECOMMENDED	EMBEDDED SYSTEMS
Processor	Dual-core 2,3GHz	Quad-core 2,7GHz or faster	NVIDIA® Jetson Nano, TX2, Xavier, Orin
RAM	4GB	8GB	8GB
Graphics Card	NVIDIA® GPU*	GTX1060 or higher**	NVIDIA® Jetson embedded GPU
Connectivity	USB 3.0	USB 3.0	USB 3.0 / GMSL2***
Operating System	Windows 10/11, Ubuntu 20.04, 22.04		NVIDIA® Jetson Linux (L4T)

Figure 2.3 Recommended laptop specifications to run ZED SDK software (Stereolabs Inc., 2024b).

During this thesis work, two laptops were used with their specifications summarized in Table 2.6. This includes the stages for which each laptop was used.

Table 2.6 Laptop types and specifications used throughout thesis work for running training and tests.

	Usage	GPU	Operating System
Laptop 1 **Asus TUF 15**	Training Field Test 1 Field Test 3	NVIDIA GeForce RTX 4050	Windows
Laptop 2 **Asus Zenbook Pro 14**	Field Test 2	NVIDIA GeForce RTX 4060	Windows

2.2.4   Drone

A drone was used at each of the test sites to capture orthophotos and enable the positional extraction of points of interest, such as tree locations. The DJI Mavic 3E was used, with specifications outlined in Table 2.7 (DJI Enterprise, 2022).

Table 2.7 Drone model and specifications

	Sensor	Field of View	Aperture	Focus
DJI Mavic 3E	4/3 CMOS, effective pixels: 20 MP	84°/24mm	f/2.8-f/11	1 m to ∞

2.2.5 Vertex Laser Geo

The Vertex Laser Geo (Fig. 2.4) is a high-precision laser rangefinder with a built-in GPS and compass. It allows the measurement and mapping of forested areas through consecutive point-and-shoot distance and position measurements. This includes an internal GPS of accuracy 2.5 m in open terrain and distance ranging up to 700 m with an accuracy of 4 cm (Haglof Sweden, n.d.). This tool was used for tree positioning during Test 3.

Figure 2.4 Vertex Laser Geo

2.3 Additional Material

A caliper (Fig. 2.5) was used to measure tree diameter at breast height (DBH).

Figure 2.5 Tree caliper (Dendrotik, 2016).

Stakes mounted with paper identifiers (Fig. 2.6) were used to mark control points. Some were used as control points for the aerial drone flight, while others were markers for the start and end positions of test paths that will be run using the stereo camera.
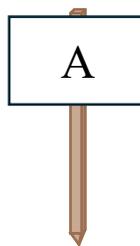
Figure 2.6 Stake position markers

3.0 Methods

This section will outline the phases of data collection and processing, including preliminary work to get equipment and software running, preliminary surveys of the study areas, the two tests in the poplar test site, and one test in the forest test site. Field survey sections will include discussion of environmental considerations will be taken into account for the preparation and execution of any on-site work in this study area in order to ensure minimal impact and damage to the site. This includes the use of non-damaging control point or tree markers, no waste items or materials left behind, and the mitigation of any other behaviours that may prove harmful to the site. Potential safety hazards will also be identified on site to ensure that hazardous areas or actions are avoided. This section will also outline the methodology for developing the processing procedures. This includes the calculations used and determined order of data-processing steps, as well as any significant considerations or assumptions taken into account during this process.

3.1 Study Areas

There are two study areas used for the development of this thesis (Fig. 3.1). The first is a poplar plantation in Legnaro, PD. (see section 3.1.1 for details) with organised rows of trees and minimal other features present that would typically be found in a more natural forest setting, such as other vegetation or boulders. This will simplify the environment that the stereo camera must take in for preliminary training and testing of the equipment. The second study area is in a mountain forest in Taibon Agordino, BL. (see section 3.1.2 for additional details) with a more natural forest structure. This will be used to test the stereo camera in a more complex "natural" forest state so comparison can be made of the stereo camera's capabilities depending on the environment it is taking in.
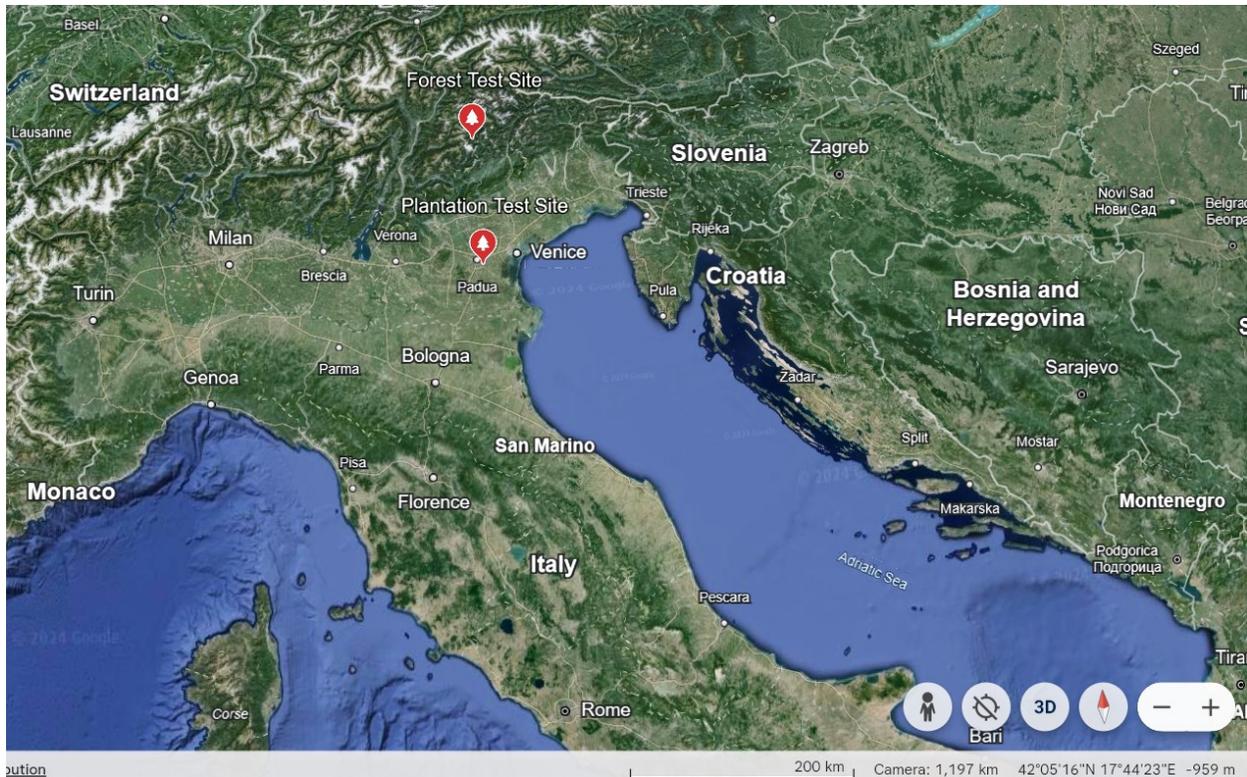
Figure 3.1 Plantation Test Site and Forest Test Site locations in Northern Italy

3.1.1 Plantation Test Area

The controlled test area will take place on a privately owned poplar plantation in Legnaro, PD (Fig. 3.2), with multiple advantages to its location and features. It is positioned 2.5 km from the University of Padova - Campus Agripolis (research base) and is 2.19 ha in area (Fig. 3.3). Key features of this site making it ideal for preliminary testing include organized rows of trees, making for easy distinction of how spacing and organisation of identified trees should appear through data collection. It also hosts a lack of other various "objects" that may interfere with preliminary equipment calibration and testing, including other vegetation, dead wood, rocks, etc. that may be present and unorganised in a more natural setting.
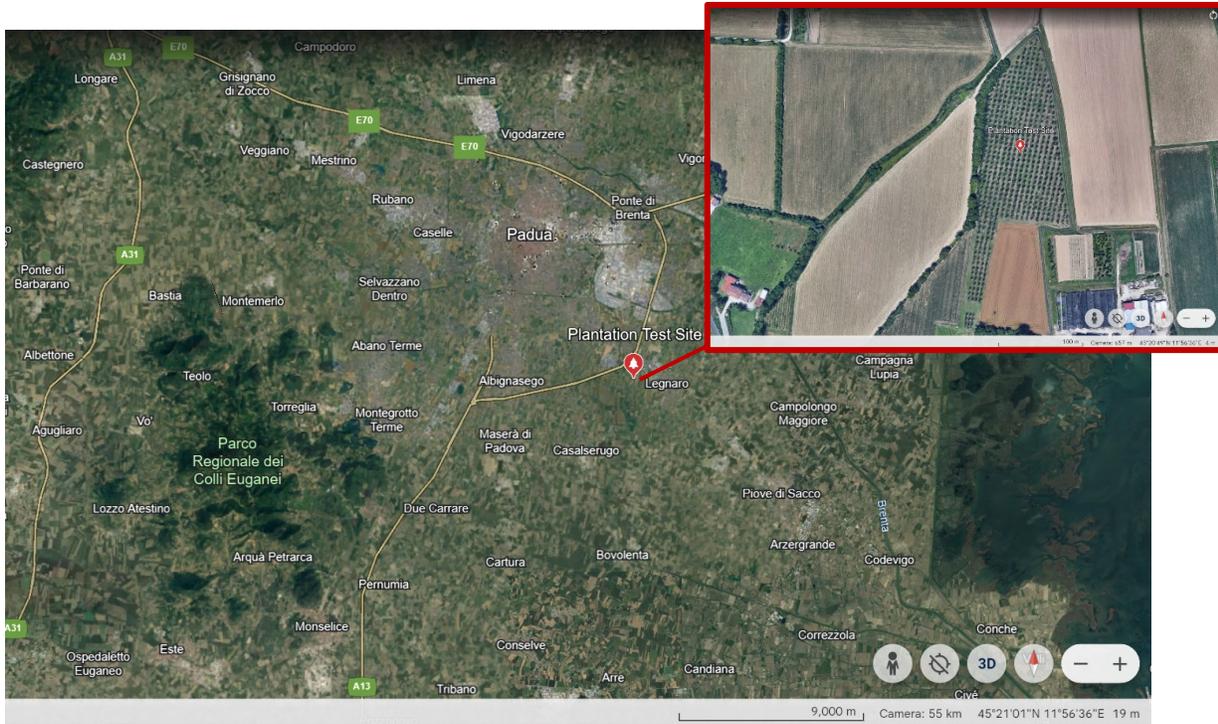
Figure 3.2 Legnaro plantation test site general location
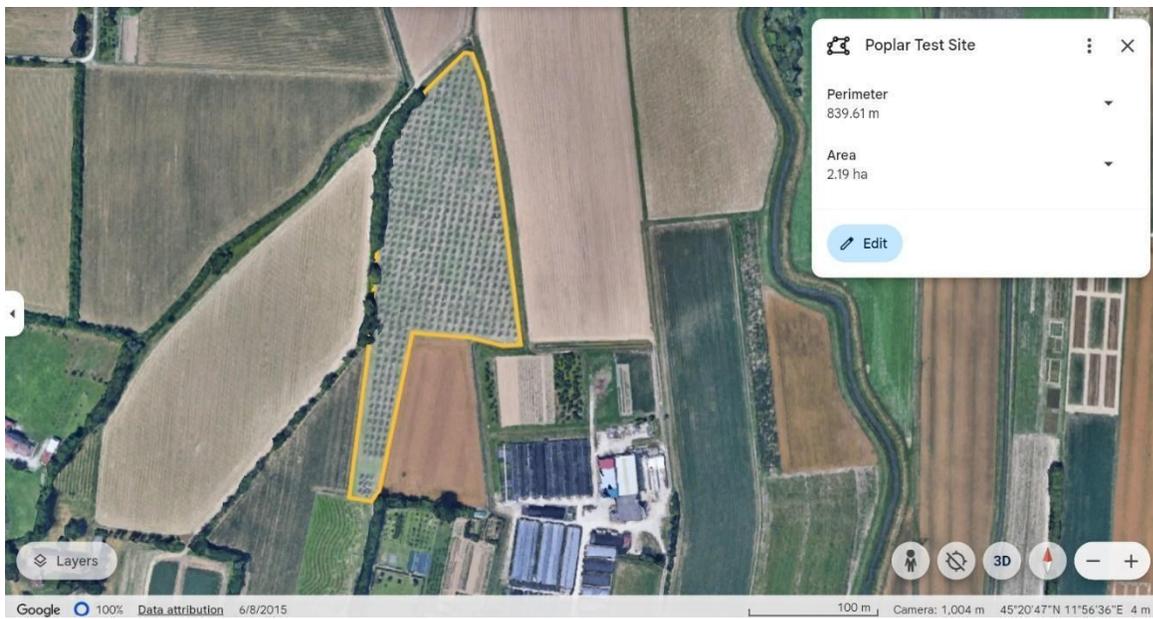
Figure 3.3 Legnaro plantation test site magnified location with special attributes

3.1.2 Natural Forest Test Area

The natural forested area test will take place on a privately owned forest road in Taibon Agordino, the province of Belluno (Fig. 3.4). It is 0.26 ha in area (Fig. 3.5), and key features of this site include the presence of many unorganised trees, as well as other vegetation, dead wood, rocks, etc. that present a more complex, natural environmental landscape.
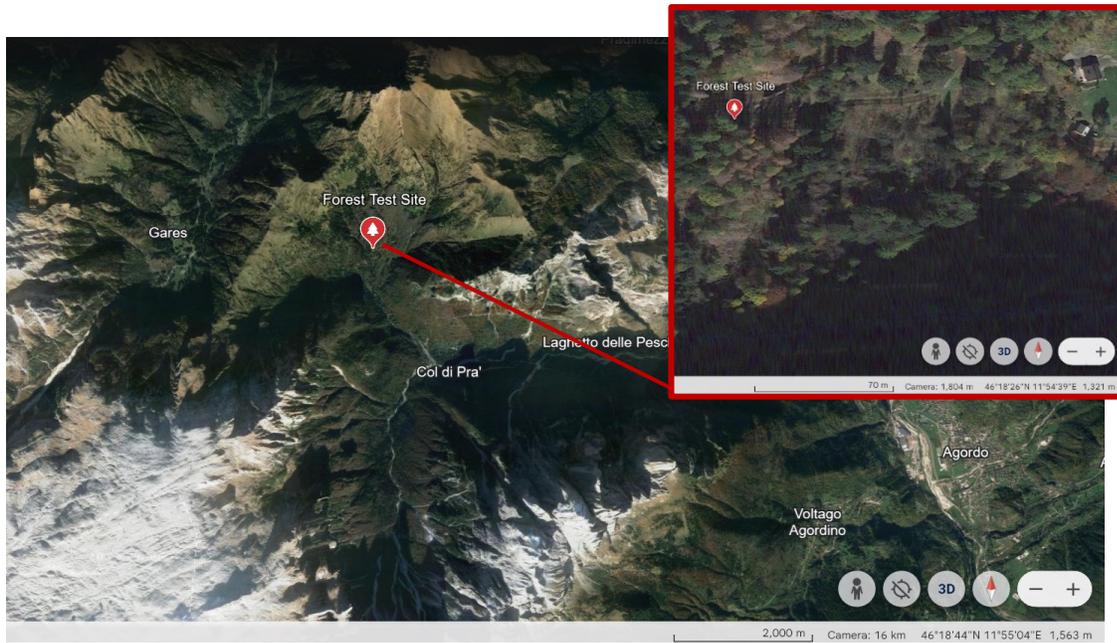
Figure 3.4 Forest test site general location
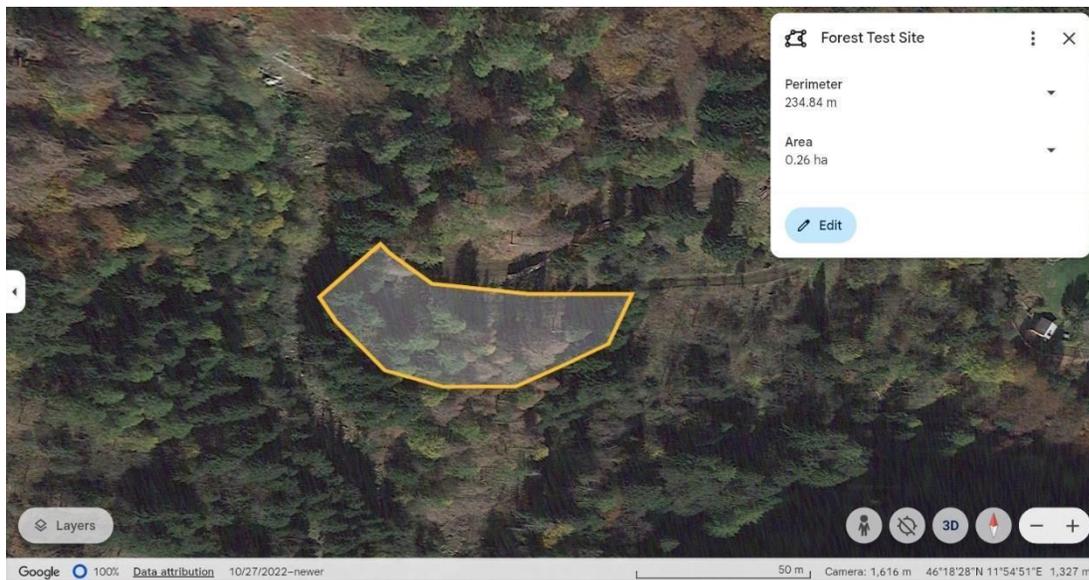
Figure 3.5 Forest test site magnified location with special attributes

3.2 Study Preparations

3.2.1   Dataset Preparation and Training

Preparations commence with the instillation of software programs for the ZED 2i stereo camera to run and log data. This includes Python, ZED SDK software, OpenCV, and PyTorch.

Once equipment and software are acquired, the software training for successful implementation with the stereo camera must be completed. As trees are the primary point of interest for preliminary camera testing, software training will commence with tree trunks. Implementation of tree detection using YOLOv8 was done using open-source tree datasets from Roboflow (roboflow, 2023a) (roboflow, 2023b). Trees in the test area (also common in the surrounding natural forest areas where this type of technology may be implemented) are namely *populus*, deciduous trees, and therefore the height of observation that is likely to be used in the camera's field of vision for object detection, identification, and measuring, will be namely sections of tree trunk. Therefore, the dataset chosen for the training of the stereo camera was filtered to trees with bounding boxes around their defining features such as their trunks so it will learn to identify these features in its test setting. The refined dataset used was comprised of 1,291 images in .jpg format and 3,521 annotations in .txt format, ensuring each image had corresponding annotation files to identify the positions of each object in each photo for model training.

The dataset was then separated into three sections with 70% of images and annotations organised into *Train*, 10% into *Test*, and 20% into *Valid*. A .yaml file was configured to specify the dataset structure for YOLO, and then the model could be trained using yolov8n.pt as the base with image size (imgsz) set to 640x640, 125 epochs, a batch size of 16, and 4 workers. Training (Fig. 3.6) was conducted on Laptop 1 (materials Section 2.2.3), for which the GPU had 6,141 MiB of memory, with 243 MiB in use during training. The GPU utilization was at 16%, operating at a temperature of 41°C, and the CUDA version was 12.3. The trained weights (.pt format) were then used in the developed Python code using ZED SDK software, OpenCV, and PyTorch. This took 5-6 hours, but can vary based on the laptop in use.

Using this weight file, the stereo camera is then able to pair with this trained model to detect trees and record the necessary information about each identified observation. This information is

set to record in a CSV file. A secondary CSV file will also record each detected object with a single record of only when it was closest to the camera's position.

Figure 3.6 Sample of training stereo camera to identify tree trunks, showing bounding box and accuracy of prediction of the model

In addition to the interface setup for the stereo camera, logging systems must be established for the independent GNSS system so that they may run consecutively. GPS recordings were facilitated by connecting the EMLID Reach M2 to the laptop via USB and recording the GPS data in a CSV file.

3.2.2 Planning Test Paths

Three different test path types will be introduced, with varying levels of complexity in direction of motion. They are anticipated to introduce increasing complexity of environmental mapping for the stereo camera, with unpredictable movement or increased chance of repetition of observation of same objects.

The first is a linear path (Fig. 3.2a), which will move at a steady, moderate pace through a selected row of trees. This simulates the most basic and straightforward test run in a forested setting and is anticipated to garner the most clear and accurate results that can be used to compare to more complex test routes.

The second is a random "snake" path (Fig. 3.7b) and is a closer simulation to real-life anticipated forestry applications in which movement will not be in a constant straight direction.

The third is a spiral path (Fig. 3.7b) and anticipated to introduce the most complex data collection scenario as it offers the most change in direction and much higher object repetition.
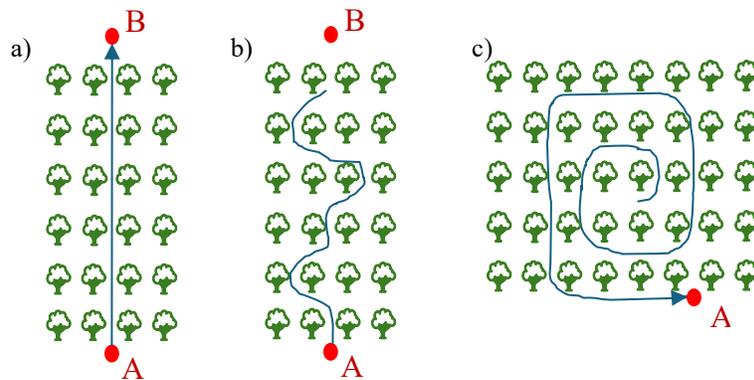
Figure 3.7 Three types of planned test routes, a) Linear, b) Snake, and c) Spiral

These varying route types will be taken into account in the data analysis and comparison of results to conclude which elements impact the accuracy and precision of object detection, identification, and sizing.

3.3 Field Survey 1

Field survey 1 was conducted in the identified test area populus plantation (Fig. 3.8). The expectations for this survey include the successful collection of object detection data for each of the three outlined test path types, including varying speeds, to identify and classify relevant objects (trees) throughout various times of the day.

The date, weather conditions, and environmental conditions for the day of this survey are as follows:

- Date: February 21, 2024
- Weather: 8-15°C, sunny, low wind.
- Tree state: bare, no blooming of leaves yet

Figure 3.8 Field Survey 1 test site conditions

Hazards identified on site include a steep ditch bordering the south-east perimeter of the plantation (Fig. 3.9a). Beekeeping is also conducted in the immediate area (Figure 3.9b), so due diligence will be used to ensure there is no disruption to their bee colonies, nor is anyone close enough that additional safety measures (any personal protective equipment) would be necessary. In general, caution will be taken when executing any tests, as focus on steadying and properly positioning equipment may increase risk of trips and falls.

Figure 3.9 Hazards on plantation test site a) Ditch and b) Beekeeping

3.3.1 Aerial Survey

An aerial survey can be beneficial as an overview of the survey area, as well as a control survey. It enables the positional comparison of data collected using the stereo camera vs the positions calculated using the obtained and adjusted orthophotos using high resolution imaging.

The base receiver was positioned on the eastern side of the plantation with minimal coverage or obstruction. Control points on the ground were marked, and their position obtained using the RTK GNSS receiver in order to be able to georeference those points in the orthophotos. Drone control points A, B, C, and D were set (Fig. 3.10).

Figure 3.10 Drone control points

The drone flight and processing details are as follows:

- Drone model: DJI Mavic 3E
- Number of photos taken: 377
- Processing software: Agisoft Metashape Professional 2.1.1

All depth maps were processed with moderate or high quality filtering settings, and the point cloud was generated with high quality setting for constructing an orthomosaic with 1.1cm/pix resolution.

Figure 3.11 Drone orthophoto of plantation test site

3.3.2   Control Survey

To execute the planned test route types, three control points (A, B, C) were set (Fig. 3.12). These control points with "known" (GPS measured) coordinates act as the start and end points to the varying test routes executed.

Figure 3.12 Plantation test route control points

For a control data set, the position of trees from two rows of trees running parallel to the forecasted paths A to B and B to C were measured using the rover (Fig. 3.13). These are anticipated to be the trees within reasonable field of view and distance from the camera as it makes its way along the anticipated routes. This will be the positive control, providing comparable positions once positions of trees identified by the stereo camera are calculated by converting their local coordinates.

Figure 3.13 Tree position control survey using Emlid Reach M2 rover

Following the same projected paths, manual measurements of the diameters of the same two rows of trees running parallel to the forecasted paths were documented. This was done using a caliper at DBH, documenting the diameter of both the south trunk side of the tree and east trunk side of the tree (Fig. 3.14).

Figure 3.14 Tree diameter measurements using caliper

It should be noted that while part of the surveying procedures, the diameter measurements obtained for comparison with dimensioning performed by the stereo camera will not be compared or analysed in comparison to dimensioning obtained by the stereo camera within the scope of this thesis.

3.3.3 Test Preparation

Before logging any data, it is ensured the rover and ZED 2i Stereo Camera are correctly running and able to log data. A safety check can also be performed to ensure that the test area is safe to navigate. This includes identifying any hazards, that equipment is safely set up and anything mounted is secure.

By mounting the rover on the camera (Fig. 3.15), the position tracked with this will be used as the zero coordinates for each object identification made from the stereo camera (the stereo camera will only provide a local coordinate position based on the zero position). These two surveying results will be used in combination to convert the local coordinates of the identified objects to objects with global coordinates (WGS84, UTM 32N).

Figure 3.15 Stereo Camera with rover centrally mounted on top

3.3.4 Data Collection

For this test, a combination of Laptop 1 and Stereo Camera 1 was used. The camera is held at chest height, level, and slightly extended from the body so as not to provide additional interference with satellite signals.

With the camera positioned over the designated starting control point, programs for stereo camera and GPS data collection can be started and movement from the start point to the end point

may begin. Movement is held at a steady pace the running codes as and video capturing process can be monitored to ensure not only that the camera appears to be held properly and the path being walked is linearly, but also that the tree observation and the corresponding data presentation are functioning correctly to capture observation data. Once the camera arrives over the end point of the test path, all data logging software is suspended, and it can be checked that the collected data has been correctly stored.

This process is repeated for all necessary test paths, after which the camera and GPS can be safely disconnected from the laptop. It is noted that an important consideration throughout all processing procedures and data analysis is the time discrepancy between any laptop being used, and any external hardware logging data.

Table 3.1 is a summary of all test routes taken, their start and end points, timestamps, speed, and the number of tree observations that were made.

Table 3.1 Test 1 summary of test route observation data

	Start Pt.	End Pt.	Path Type	Start Time	End Time	Speed	# Observations
1	A	B	Linear	11:32:58.60	11:34:38.60	Normal	1611
2	B	A	Linear	11:17:38.60	11:19:35.60	Normal	2263
3	-	A	Spiral	14:56:41.60	15:00:36.60	Normal	2814
4	C	A	Snake	14:50:09:60	14:51:58.60	Normal	1608
5	B	C	Linear	12:10:22.60	12:11:53.60	Normal	767
6	B	C	Linear	15:14:05.60	15:15:15.60	Normal	1243
7	C	B	Linear	12:15:25.60	12:16:32.60	Normal	587
8	C	B	Linear	15:17:43.60	15:18:47.60	Normal	973
9	A	B	Linear	12:05:16.60	12:06:16.60	Fast	922

For the purpose of this test, "normal" pace depicts a slower steady pace. Averaging the duration of tests between points A and B (~67m apart) normal pace is approximated to be 0.6m/s. "Fast" depicts an average or slightly more accelerated walking pace. Averaging the duration of fast tests between points A and B, fast pace is approximated to be 1.2m/s.

3.4 Processing Procedures 1

Processing the first round of test data will prompt the initial development of methodology procedures for processing and calculating informative data collected from a stereo camera survey. Such procedures will be expressed in order of which they were devised. All data amalgamation and calculations were done in Excel, while all data visualization and representation is in QGIS.

3.4.1   Timestamp Correction, Refinement, Matching

To render GNSS and stereo camera data comparable, the timestamps of both datasets must be adjusted. The GNSS data from Test 1 is 1h behind, therefore, 1h must be added to GNSS timestamps. The stereo camera logged time is logged based on the interface clock, so in this circumstance that is the computer clock. By comparing timestamps of overlapping stereo camera positions (logged once per second) and GNSS positions (logged five times per second) this is determined to be 18.6 seconds behind the GNSS time (Fig. 3.16). Therefore, 18.6 seconds must be added to the stereo camera timestamps.
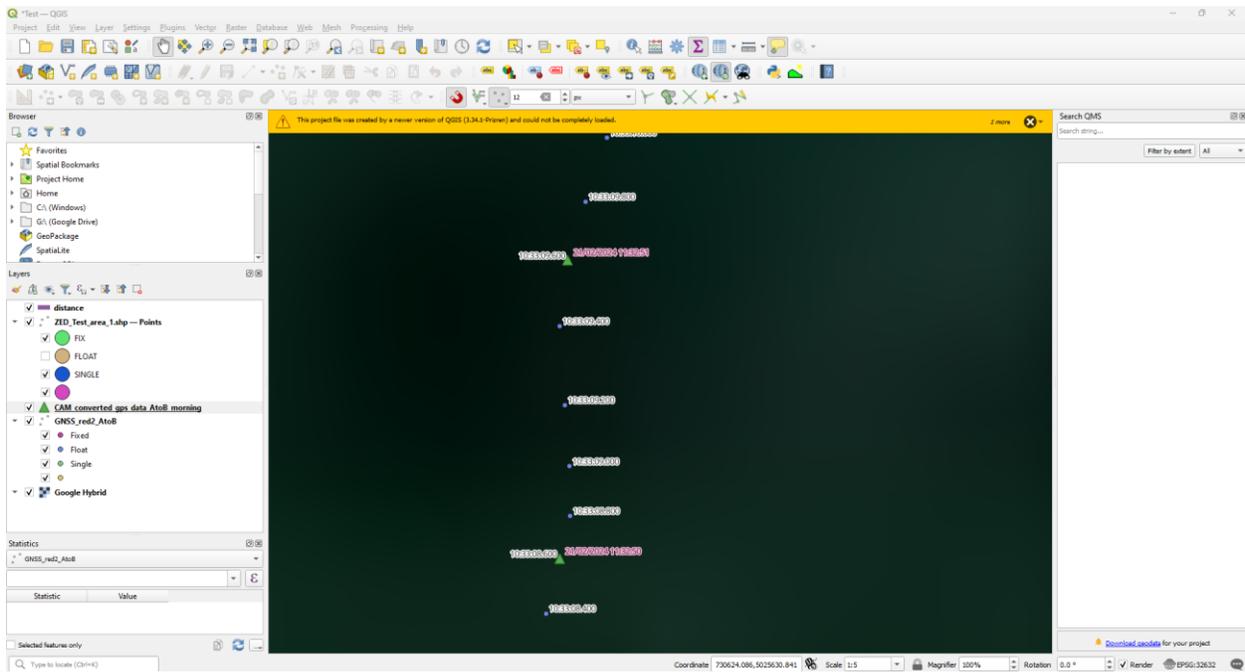
Figure 3.16 Test 1 GNSS vs stereo camera data showing time discrepancy between matching position points once per second

As the rover ran continuously and independent from the stereo camera test runs, it is important to determine the corrected start and end time of the stereo camera run time for each test conducted in order to trim the corresponding matching GNSS data. This is obtained by looking at the first and last timestamp of the corrected stereo camera timestamps for each test. The GNSS data during this time range can then be isolated as the identified path taken during an individual test run. Key GNSS data to take into consideration from the output files of the rover include the adjusted timestamps of observation positions, and the positional data of those positions (specifically longitude and latitude).

28

From the stereo camera test data, important data to isolate includes the adjusted timestamp of each observation, as well as their corresponding *Object ID*, *Position [X, Y, Z] Confidence*, and *Distance*. In amalgamating these two sets of data, the objective is to use the GNSS data for the latitude and longitudinal position at each observation point along the test path. Then, matching timestamps, the stereo camera data that detected objects in a local coordinate system can be used in combination to calculate the global position of each object detected at each observing position along the test path. This is presented in Figure 3.17, with blue representing the local coordinate system and green representing the positioning of that local coordinate system in the global coordinate system.
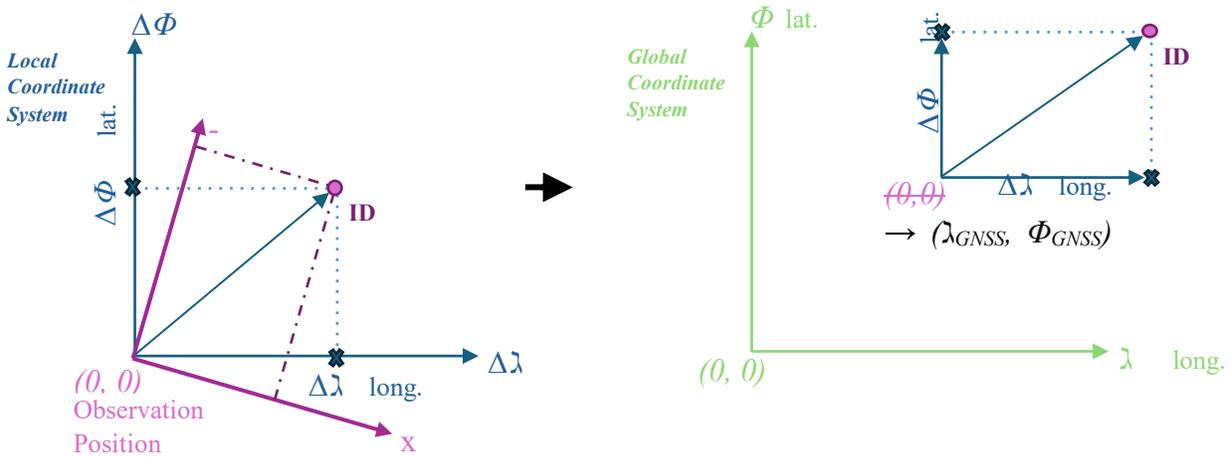
Figure 3.17 Calculation of global position of an observation made in a local coordinate system, given the global position of the zero-position of the camera

3.4.2   Global Position Calculation

The first value needed for this conversion is the adjustment rotation angle. This is based off the azimuth calculated according to the north direction of the local coordinate system. In this case, that is -*Z*. To determine the direction of the camera at any given observation point, an assumption is made that the carrying of the stereo camera is exactly straight, as the IMU data will not be incorporated in this preliminary calculation. Therefore, to get the direction, the vector between one observation point and the average positional direction of the next five observation points is taken as the "true" positional direction. The recording frequency is 5Hz, therefore that will be 5 positions over the course of one second. Given this assumption, the azimuth of that line vs true north of the global coordinate system can be calculated.

Given that not all tests are conducted in the same direction with directions falling in the same directional quadrant, two main variations to the azimuth calculation are prevalent, Eq. 3.1 and Eq. 3.2 (where if the azimuth will land in quadrant I then Eq. 3.1 is used, otherwise the azimuth is calculated with Eq. 3.2) (Fig. 3.18).

Eq. 3.1 Azimuth Equation (Quadrant I)

$$\alpha = 90 - \theta$$

Eq. 3.2 Azimuth Equation (Quadrant II, III, IV)

$$\alpha = 450 - \theta$$
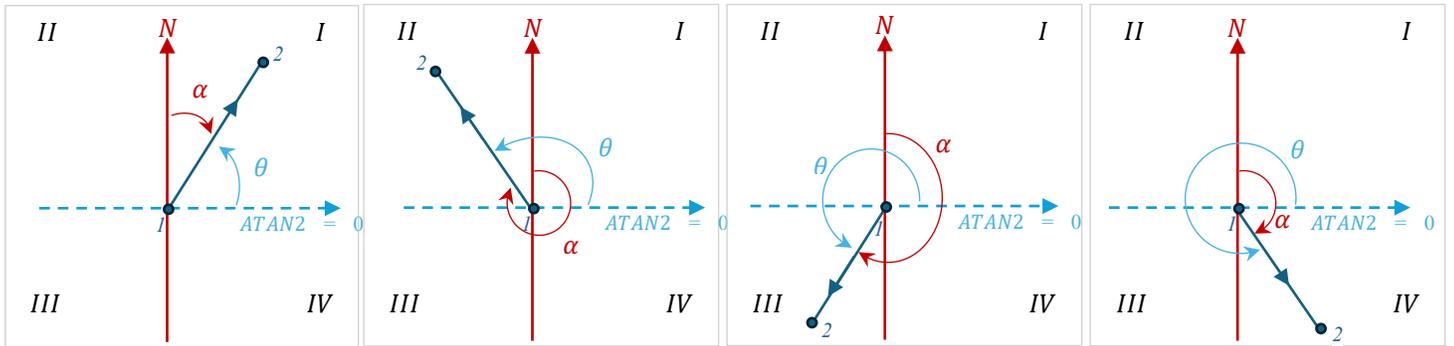
Figure 3.18 Azimuth equation based on directional quadrant

To calculate the angle of interest, theta, given two components x and y, the sign of these components varies the equation. (Veness, C., 2022; MedCalc Software Ltd, 2024; Omni Calculator, 2024)

Eq. 3.3 ARCTAN2 Equation for azimuth calculation

$$\theta = ARCTAN2(x, y) \quad where \quad x = sin\Delta\lambda cos\phi_2$$

$$y = cos\phi_1 sin\phi_2 - sin\phi_1 cos\phi_2 cos\Delta\lambda$$

$$and\ ARCTAN2(x,y) = arctan\left(\frac{y}{x}\right) \qquad if\ x > 0$$

$$= arctan\left(\frac{y}{x}\right) + \pi \qquad if\ x < 0\ and\ y \geq 0$$

$$= arctan\left(\frac{y}{x}\right) - \pi \qquad if\ x < 0\ and\ y < 0$$

$$= +\frac{\pi}{2} \qquad if\ x = 0\ and\ y > 0$$

$$= -\frac{\pi}{2} \qquad if\ x = 0\ and\ y < 0$$

$$= undefined \qquad if\ x = 0\ and\ y = 0$$

Given the azimuth, the rotation angle can be calculated (Eq. 3.4) (Fig. 3.19). Given this rotation angle and local position of a given observation, the change in longitude and latitude from the camera position to the observation position can be calculated (Eq. 3.5).
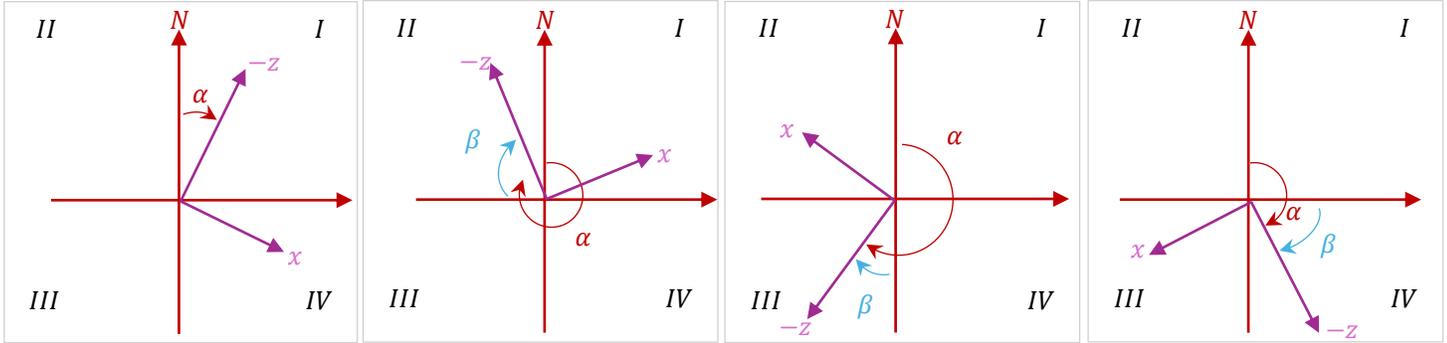
Figure 3.19 Rotation angle calculation based on directional quadrant

Eq. 3.4 Rotation angle calculation based on directional quadrant

if $0° < \alpha < 90°$ if $270° < \alpha < 360°$ if $180° < \alpha < 270°$ if $90° < \alpha < 180°$

$\beta = \alpha$ $\beta = \alpha - 270$ $\beta = \alpha - 180$ $\beta = \alpha - 90$

Eq. 3.5 Change in longitude and latitude equations based on directional quadrant

$\Delta\Phi$ lat = (-Z) cos (β) - (X) sin (β) $\Delta\Phi$ lat. = (-Z) sin (β) + (X) cos (β) $\Delta\Phi$ lat = -((-Z) cos (β) - (X) sin (β)) $\Delta\Phi$ lat. = -((-Z) sin (β) + (X) cos (β))

$\Delta\lambda$ long. = (-Z) sin (β) + (X) cos (β) $\Delta\lambda$ long = -((-Z) cos (β) - (X) sin (β)) $\Delta\lambda$ long. = -((-Z) sin (β) + (X) cos (β)) $\Delta\lambda$ long = (-Z) cos (β) - (X) sin (β)

Given this change in longitude and latitude from the camera position to an observed tree, the observation's global position is obtained by adding the change in longitude to the corresponding timestamp's camera position longitude, and the change in latitude to the camera position latitude (Eq. 3.6).

Eq. 3.6 Final global positions (longitude and latitude) of a given observation

$$\Phi Lat\_observation = \Phi lat\_cam + \Delta\Phi\ lat$$

$$\lambda Long\_observation = \lambda long\_cam + \Delta\lambda\ long$$

3.5 Field Survey 2

Field survey 2 was conducted in the identified test area populus plantation (Fig. 3.20). The expectations for this survey include the successful collection of object detection data for each of the three outlined test path types, including varying speeds, to identify and classify relevant objects (trees) now that the leaves have come in on the tree canopy.

The date, weather conditions, and environmental conditions for the day of this survey are as follows:

- Date: May 13, 2024
- Weather: 20-22°C, sunny, some clouds, low wind.
- Tree state: full foliage

Figure 3.20 Field Survey 2 test site conditions

3.5.1 Aerial Survey

Despite change in foliage presentation, the tree positions are assumed to be the same as before, and thus no new aerial survey was conducted. Therefore, for the purpose of data comparison with data extracted form an orthophoto of the site, the orthophoto obtained during Field Survey 1 will be sufficient.

3.5.2 Control Survey

The same path control points (A, B, and C) are used for this second field survey. Additionally, all other control elements remain the same including GNSS tree positions taken with GPS, the planned paths for different test styles of routes through the trees and the manual tree data measured by caliper for the trees in proximity of interest. The only difference for this test is the

position of the base. The base is positioned on the southern end of the test site for this survey (Fig. 3.21).

Figure 3.21 Control points during Field Survey 2, with new GNSS base position

3.5.3   Test Preparation

Preparation for carrying out the test procedures will remain the same as during Field Survey 1, referenced in section 3.3.3.

3.5.4   Data Collection

The carrying out of the test procedures will remain the same as during Field Survey 1, referenced in section 3.3.4. Table 3.2 is a summary of all test routes taken, their start and end points, timestamps, speed, and the number of tree observations that were made. It should be noted that for test 8, the video was not recorded. All other observation data is still available.

Table 3.2 Field Survey 2 summary of test route observation data

	Start Pt.	End Pt.	Path Type	Start Time	End Time	Speed	# Observations
1	A	B	Linear	11:13:17.80	11:15:01.60	Normal	1611
2	B	A	Linear	10:47:57.40	10:49:38.80	Normal	2263
3	-	A	Spiral	10:56:16.60	11:00:48:00	Normal	2814
4	C	A	Snake	11:08:18.00	11:10:26.20	Normal	1608
5	A	B	Linear	11:36:24.20	11:37:18.20	Fast	767
6	B	C	Linear	11:20:32.40	11:22:00.60	Normal	1243
7	C	B	Linear	11:24:55.40	11:26:10.00	Normal	587
8	A	B	Linear	10:34:27.80	10:36:02.80	Normal	973

3.6 Processing Procedures 2

Processing the second round of test data will follow the established methodology procedures for processing and calculating informative data collected from a stereo camera survey. All data amalgamation and calculations were done in Excel, while all data visualization and representation is in QGIS.

3.6.1 Timestamp Correction, Refinement, Matching

The GNSS data from Test 2 is 2h behind due to daylight savings, therefore, 2h must be added to GNSS time. A different computer was used during Field Test 2, and this was determined to be 19.4 seconds behind the GNSS time (Fig. 3.22). Therefore, 19.4 seconds must be added to the stereo camera timestamps.
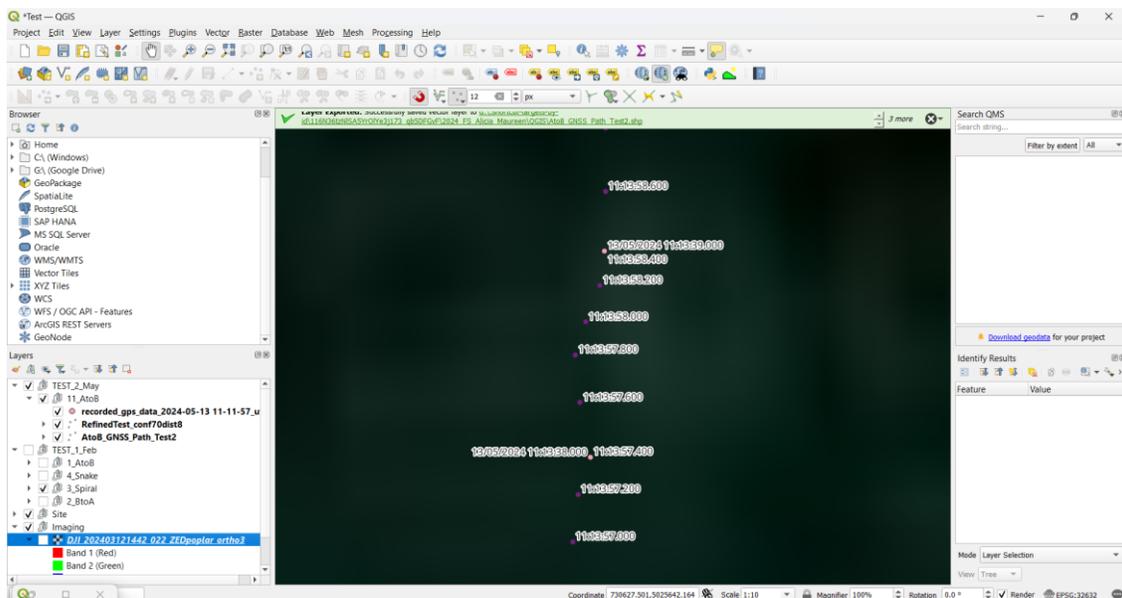
Figure 3.22 Test 2 GNSS vs stereo camera data showing time discrepancy between matching position points once per second

3.6.2 Global Position Calculation

The remaining calculations follow those outlined in *Processing Procedures 1*, equations 3.1 to 3.6. The only additional consideration was that camera observations this test were recorded to the nearest millisecond. As GPS positions only had timestamps up to 1 decimal place, in increments of 0.2 seconds, the timestamps of all camera observations had to be rounded to the nearest 0.2 second.

3.7 Field Survey 3

Field survey 3 was conducted in the identified natural forest test area (Fig. 3.23, Fig. 3.24a, Fig. 3.24b). The expectations for this survey include the successful collection of object detection data the most simplified test path type. The focus of this survey in "real world" terrain will be the comparison data collected during this test to that collected during Tests 1 and 2.

The date, weather conditions, and environmental conditions for the day of this survey are as follows:

- Date: June 13, 2024
- Weather: 15-20°C, cloudy, some rain, low wind.
- Tree state: full foliage

Figure 3.23 Field Survey 3 test site overview
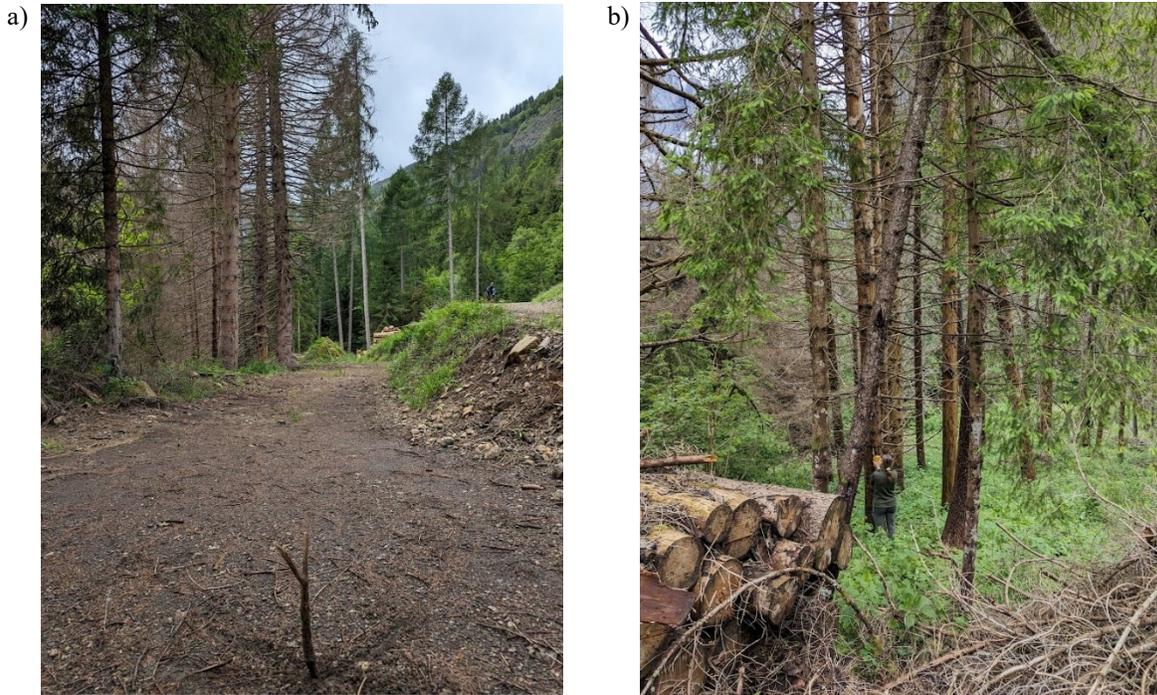
Figure 3.24 Field Survey 3 test site conditions a) along path and b) the sides of the path to be picked up by the stereo camera

3.7.1 Aerial Survey

In contrast to the plantation test site, a control survey by way of orthophotos from drone imagery will not be used for this survey. A flight was still performed for high resolution imaging, however due to foliage and the intent of specific data to be used and compared for this "real world" survey, further extraction of tree positioning, etc. from drone data will not be performed or used for Field Survey 3.

The execution of the aerial survey was to the encompass entire test area, including all potential path routes and trees anticipated to be collected in their proximity (Fig. 3.25). The flight details are as follows:

- Drone model: DJI Mavic 3E
- Number of photos taken: 220
- Processing software: Agisoft Metashape Professional 2.1.1

Figure 3.25 Drone orthophoto of forest test site

3.7.2 Control Survey

Tests performed at this site will be reduced in comparison to Field Survey 1 and 2, limited to the most basic straight paths between two points. The rigorous terrain required greater selectivity in choosing a path from point A to point B for data collection.

To execute the planned test route types, two control points (A, B) were set (Fig. 3.26). These control points with "known" (GPS measured) coordinates act as the start and end points to the routes.

Figure 3.26 Test path control points A and B

For a control data set, the position of trees anticipated to be within reasonable field of view and distance from the camera as it makes its way along the anticipated A to B path were measured with GPS and with the Vertex Laser Geo (Fig. 3.27). This allows the measurement and mapping of forested areas through consecutive point-and-shoot distance and position measurements (Fig. 3.28). This will be the positive control, providing comparable positions once positions of trees identified by the stereo camera are calculated by converting their local coordinates. Results for locations obtained by the Vertex Laser Geo and GPS will be presented in the *Results* section.

Figure 3.27 Use of Vertex Laser Geo to survey test area
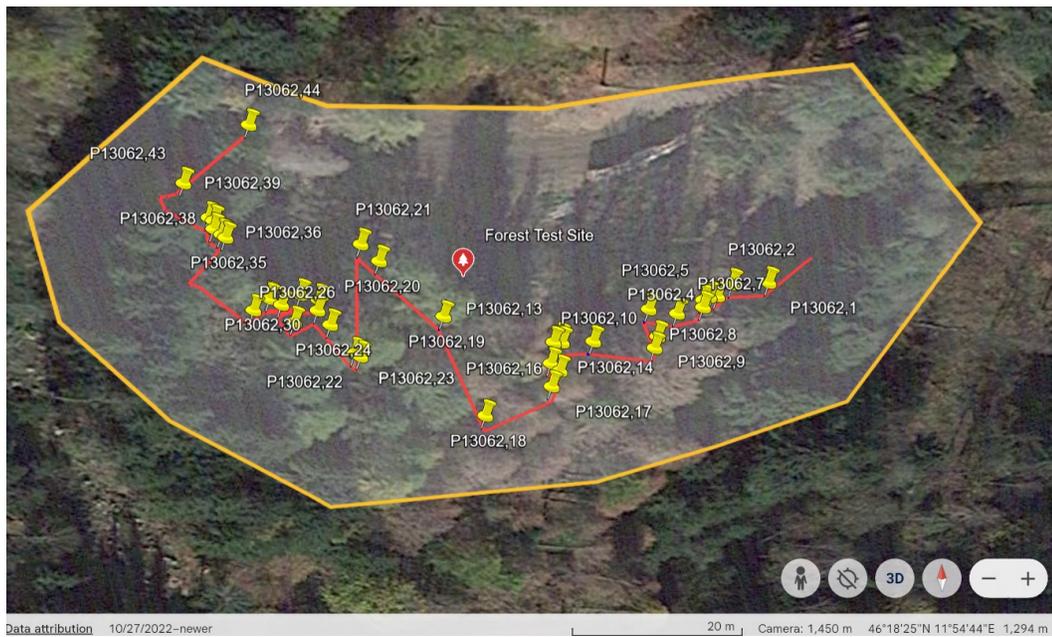
Figure 3.28 Vertex Laser Geo point and shoot path to pick up anticipated trees of interest

3.7.3 Test Preparation

Preparation for carrying out the test procedures will remain the same as during Field Survey 1 and 2, referenced in section 3.2.3.

3.7.4 Data Collection

The carrying out of the test procedures will remain the same as during Field Survey 1 and 2, referenced in section 3.3.4. Table 3.3 is a summary of all test routes taken, their start and end points, timestamps, speed, and the number of tree observations that were made.

Table 3.3 Field Survey 3 summary of test route observation data

	Start Pt.	End Pt.	Path Type	Start Time	End Time	Speed	# Observations
1	A	B	Linear	14:05:16.00	14:06:59.40	Normal	629
2	B	A	Linear	14:14:13.20	14:16:14.40	Normal	2492
3	A	B	Linear	14:18:23.60	14:20:10.60	Normal	734
4	B	A	Linear	14:26:47.00	14:28:33.20	Normal	1995

3.8 Processing Procedures 3

Processing the third round of test data will follow the established methodology procedures for processing and calculating informative data collected from a stereo camera survey. All data amalgamation and calculations were done in Excel, while all data visualization and representation is in QGIS.

3.8.1 Timestamp Correction, Refinement, Matching

The GNSS data from Test 3 is 2h behind, therefore, 2h must be added to GNSS time. Laptop 1 was used, and since the same GPS was also used as Field Survey 1, it is assumed that data from this test will also be 18.6 seconds behind the GNSS time. Therefore, 18.6 seconds must be added to the stereo camera timestamps.

3.8.2 Global Position Calculation

The remaining calculations follow those outlined in *Processing Procedures 1*, equations 3.1 to 3.6. The only additional consideration was that camera observations this test were recorded to the nearest millisecond. As GPS positions only had timestamps up to 1 decimal place, in increments of 0.2 seconds, the timestamps of all camera observations had to be rounded to the nearest 0.2 second.

4.0 Results

Results produced from each step of the outlined methodology will be presented. This will include orthophotos of the test area with obtained results superimposed, such as the fixed points set throughout the test area, paths taken when collecting data, and trees and objects identified. This will include comparisons of positions of points taken with GNSS positioning and the position of points identified using the stereo camera.

4.1 General Mapping

This section presents general site maps of the two test sites, including the Legnaro site GPS tree position points (Fig. 4.1), the Legnaro site drone tree position points (Fig. 4.2), the Taibon Agordino vertex laser geo tree positions (Fig. 4.3), and the Taibon Agordino GPS tree positions (Fig. 4.4).

Figure 4.1 Legnaro site tree positions from GPS

Figure 4.2 Legnaro site tree positions extracted from drone point cloud
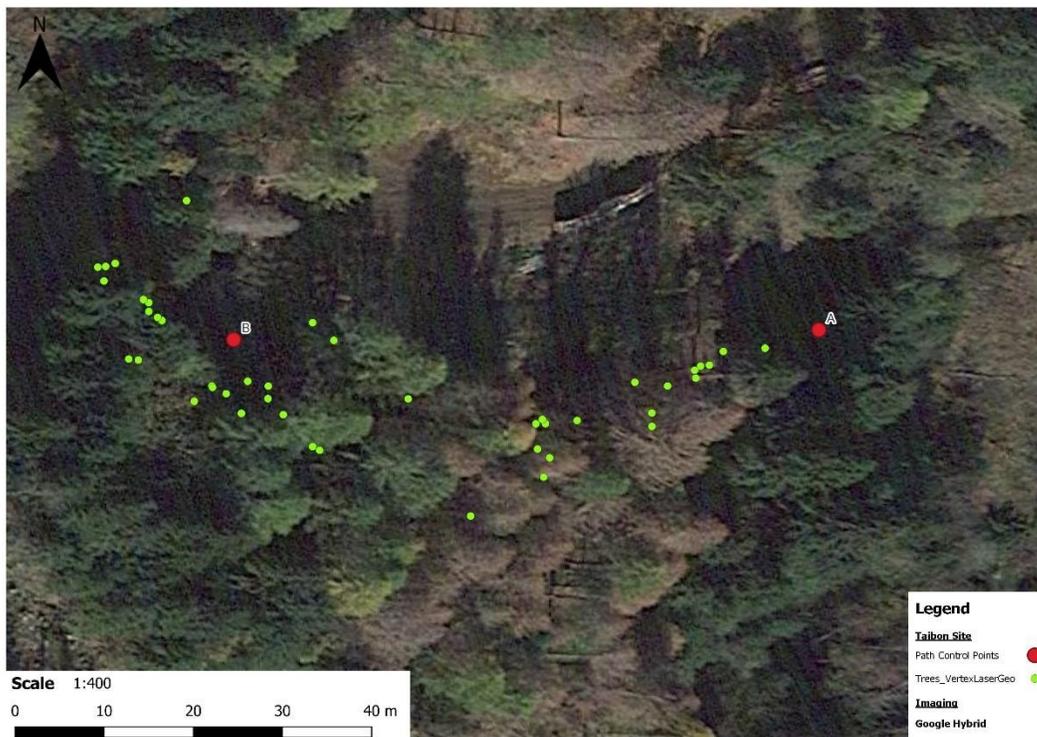
Figure 4.3 Vertex Laser Geo tree positions to use for result comparison of stereo camera observations

42

Figure 4.4 GPS tree positions to use for result comparison of stereo camera observations

Processing the Taibon site orthophoto also included the generation of a point cloud, with path segments presented in Figure 4.5 and 4.6.

Figure 4.5 Taibon point cloud looking west, Going from A to B near start of path

Figure 4.6 Taibon point cloud looking west, Going from A to B near end of path

4.2 Calculated Observation Point Global Positions

For each test path of all field tests performed, all observations and their global positions were calculated. Though this is a result from the calculation procedure, refinement of reliable data is necessary. Despite observations beginning and ending at explicit timestamps, movement along the path may not have begun exactly as the start of survey was initiated. Therefore, the GNSS positions along the path of each test route must be trimmed by examining the start and end of each test path, identifying any clustering of GNSS positions before consistent movement began in the test or at the end of the test. This leads to identifying the timestamp of the first consistent path point as well as the last consistent path point. For the last reliable timestamp, it is then necessary to identify the timestamp of 5 positions (1 second) before this last reliable point. This is because observation positions calculated from a given GNSS position uses the next 5 position points averaged to calculate an azimuth, therefore the last position which can be used to calculate observations from must have 5 more consecutive reliable position points after it. This process trims the dataset to observation positions calculated only using reliable global coordinates of camera positions.

To appropriately assess results, this vast quantity of observation points must also be refined based on its quality. This will be done using two characteristics of the observation: confidence, and distance. The confidence explicitly demonstrates the reliability of the detection, with higher confidence indicating that the detection is more likely to be correctly identified and positioned. Distance can indirectly provide insight to the anticipated reliability of a detection, as stereo camera image processing is influenced by the image disparity between the two points of view incurred during a single stereo camera observation. An increase in the distance to a detection results in a decrease in the disparity between images, which makes depth estimations more challenging and less precise (Hartley & Zisserman, 2004). To refine the dataset to test results of reasonably assumed quality, an initial filtering of confidence greater than 50, and distance within 15m of the camera position was applied for each test path. Additionally, another round of filtering was performed to explicitly highlight observations with confidence greater than 70, and distance within 8m of the camera position. Results regarding the refinement of the GNSS test paths, as well as the isolated filtered results excluding observations outside of the determined reliable range of GNSS position points are presented for tests 1, 2, and 3, in tables 4.1, 4.2, and 4.3 respectively.

Table 4.1 Test 1 refined timestamps and filtered tree observations

Test 1	Full Data			Cropped Reliable Data				Conf ≥ 50 Dist ≤ 15		Conf ≥ 70 Dist ≤ 8	
	# Obs.	Start Time	End Time	# Obs.	% total obs	Start Time	End Time	# Obs.	% total obs	# Obs.	% total obs
1_ A to B	1611	11:32:58.60	11:34:38.60	1561	96.90%	11:33:00.80	11:34:37.60	235	14.59%	19	1.18%
2_ B to A	2263	11:17:38.60	11:19:35.60	2226	98.37%	11:17:41.60	11:19:35.60	777	34.33%	115	5.08%
3_ Spiral	2814	14:56:41.60	15:00:36.60	2808	99.79%	14:56:43.60	15:00:35.60	1056	37.53%	205	7.29%
4_ CtoA snake	1608	14:50:09:60	14:51:58.60	1574	97.89%	14:50:11.80	14:51:57.60	614	38.18%	135	8.40%
5_ BtoC	767	12:10:22.60	12:11:53.60		*GPS data unreliable*			-	-	-	-
6_ BtoC evening	1243	15:14:05.60	15:15:15.60	1217	97.91%	15:14:07.40	15:15:14.60	452	36.36%	72	5.79%
7_ CtoB	587	12:15:25.60	12:16:32.60		*GPS data unreliable*			-	-	-	-
8_ CtoB evening	973	15:17:43.60	15:18:47.60	921	94.66%	15:17:45.80	15:18:46.60	281	28.88%	55	5.65%
9_ AtoB fast	922	12:05:16.60	12:06:16.60	881	95.55%	12:05:19.60	12:06:15.60	145	15.73%	12	1.30%

Table 4.2 Test 2 refined timestamps and filtered tree observations

Test 2	Full Data			Cropped Reliable Data				Conf ≥ 50 Dist ≤ 15		Conf ≥ 70 Dist ≤ 8	
	# Obs.	Start Time	End Time	# Obs.	% total obs	Start Time	End Time	# Obs.	% total obs	# Obs.	% total obs
11_ A to B	4700	11:13:17.80	11:15:01.60	3681	78.32%	11:13:30.40	11:14:59.80	1327	28.23%	236	5.02%
12_ B to A	4094	10:47:57.40	10:49:38.80	4031	98.46%	10:48:00.00	10:49:31.40	1297	31.68%	236	5.76%
13_ Spiral	15043	10:56:16.60	11:00:48:00	14133	93.95%	10:56:25.60	11:00:40.60	4770	31.71%	975	6.48%
14_ Snake	6055	11:08:18.00	11:10:26.20	5915	97.69%	11:08:21.60	11:10:14.40	1573	25.98%	337	5.57%
15_ A to B Fast	3486	11:36:24.20	11:37:18.20	3430	98.39%	11:36:26.60	11:37:18.20	1240	35.57%	211	6.05%
16_ B to C	6116	11:20:32.40	11:22:00.60	5465	89.36%	11:20:34.00	11:21:51.60	1176	19.23%	238	3.89%
17_ C to B	5946	11:24:55.40	11:26:10.00		*GPS data unreliable*			-	-	-	-
18_ A to B no vid	4235	10:34:27.80	10:36:02.80	3733	88.15%	10:34:34.00	10:36:01.80	1453	34.31%	321	7.58%

Table 4.3 Test 3 refined timestamps and filtered tree observations

Test 3	Full Data			Cropped Reliable Data				Conf ≥ 50 Dist ≤ 15		Conf ≥ 70 Dist ≤ 8	
	# Obs.	Start Time	End Time	# Obs.	% total obs	Start Time	End Time	# Obs.	% total obs	# Obs.	% total obs
21_ A to B	629	14:05:16.00	14:06:59.40	604	96.03%	14:05:16.80	14:06:58.40	216	34.34%	50	7.95%
22_ B to A	2492	14:14:13.20	14:16:14.40	2485	99.72%	14:14:13.20	14:16:03.40	998	40.05%	60	2.41%
23_ A to B two	734	14:18:23.60	14:20:10.60	722	98.37%	14:18:25.20	14:20:08.60	261	35.56%	40	5.45%
24_ B to A two	1995	14:26:47.00	14:28:33.20	1980	99.25%	14:26:48.00	14:28:31.40	544	27.27%	20	1.00%

Test paths marked "GPS data unreliable" indicates that paths were either mainly comprised of "single" points (poor positional accuracy), or simply when examined on QGIS blatantly showed a skewed path. Each refined observation set from each path was saved as a .csv file and imported into QGIS.

Figure 4.7 visually represents each path's total number of observations during the test run (primary Y axis, left) and percentage of observations that remained (secondary Y axis, right) after the refinement both of confidence greater than 50 and distance within 15m, and further refinement of confidence greater than 70 and distance within 8m. Tests that were deemed unreliable are entered with 0 observations.
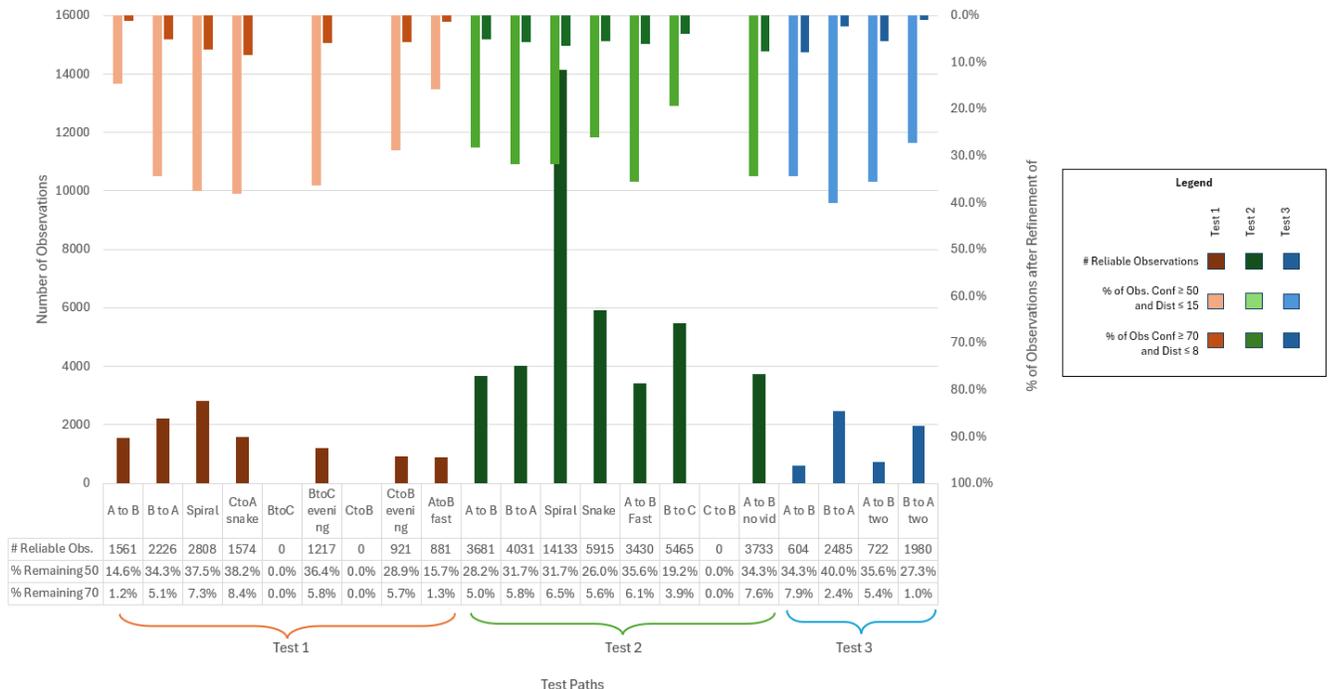
Figure 4.7 Comparison of test runs, number of observations, and % of observations maintained during stages of refinement

Figure 4.8 represents an example of these results for test path A to B during Test 2 with the filtering of 50% confidence and 15m distance range, while Figure 4.9 represents the same test path with filtering of 70% confidence and 8m distance range. Observations are labeled with their observation ID, and while the legend simply indicates observations as a singular-coloured dot, the symbology for the observations in the imaging set to "categorized" to more blatantly identify groupings of observations given same object IDs by varying the colour of different groupings of ID labels.
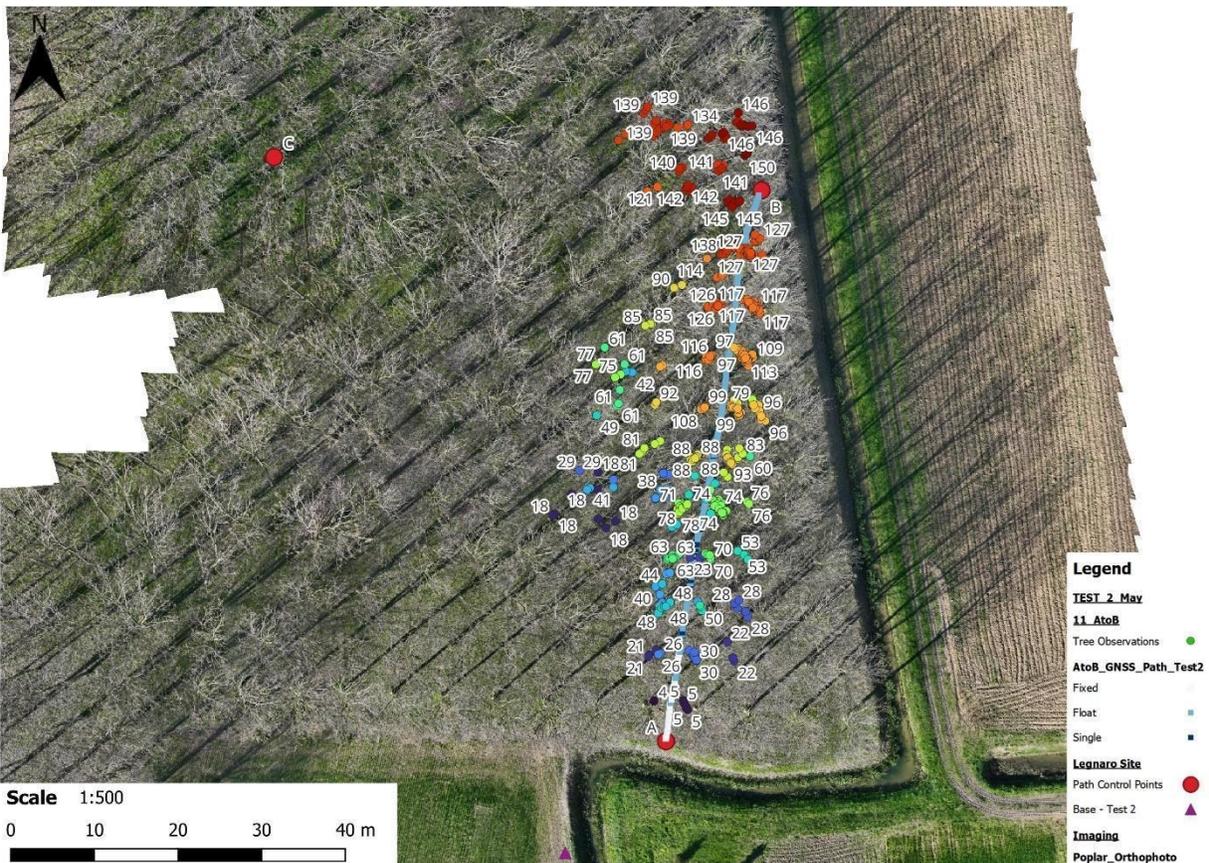
Figure 4.8 Test 2 path A to B with filtering of 50% confidence and 15m distance range
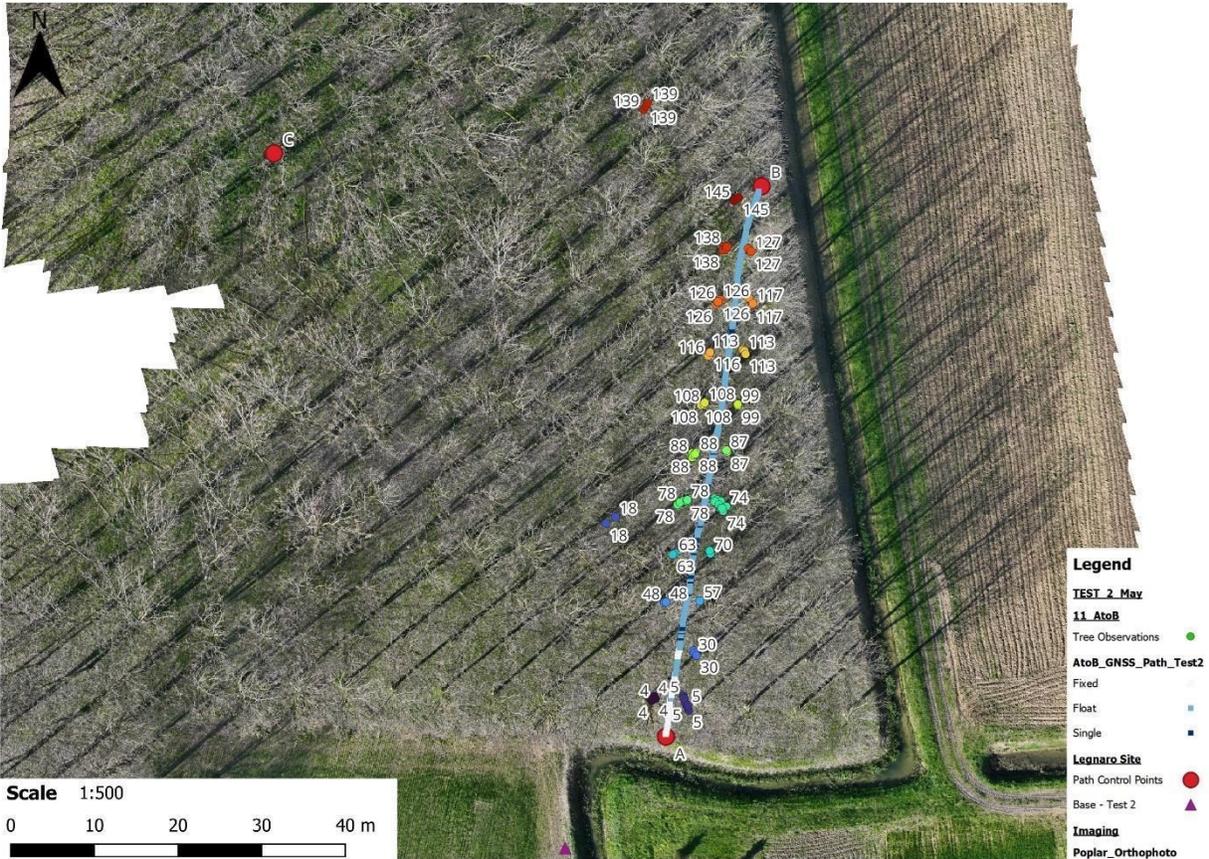
Figure 4.9 Test 2 path A to B with filtering of 50% confidence and 15m distance range

Although these results offer a range of potential tree locations, they do not provide a unified set of precise, pinpointed positions for the trees based on the various observations labeled as potential same objects. In order to achieve this, a QGIS graphical modeler was designed to take the filtered observation positions, cluster them based on proximity to each other (to simulate that they are likely referencing the same tree) and locate the centre of each cluster to pinpoint the average positions of assumed trees (Fig. 4.10).
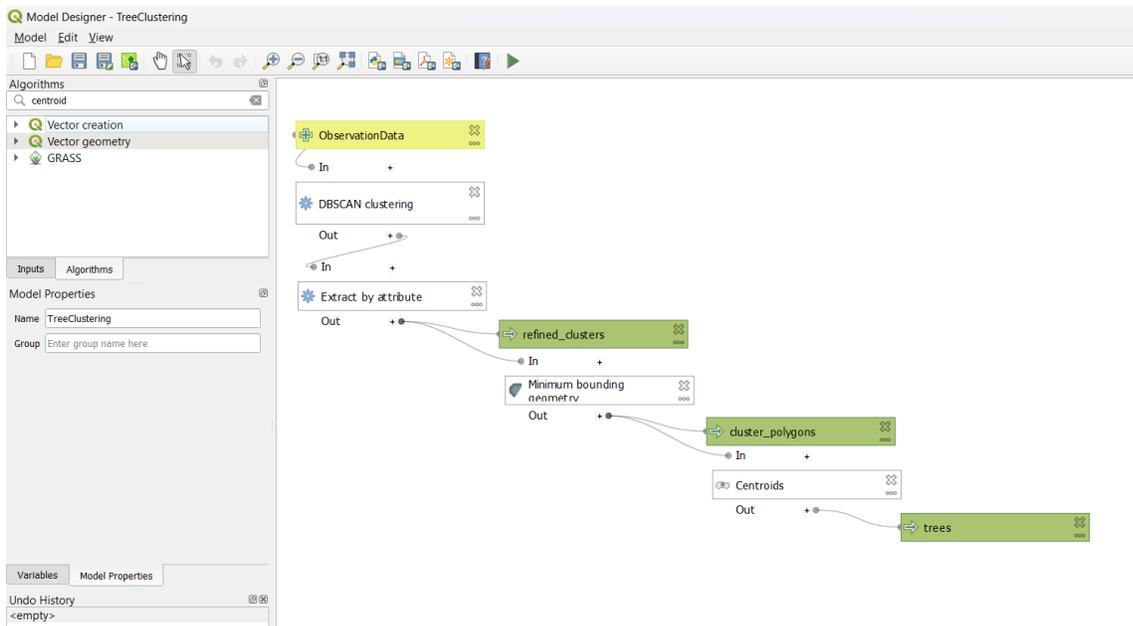
Figure 4.10 Graphical modeler to cluster and pinpoint tree positions

Clustering will take the input of a given set of refined observations, and cluster any of those observations within a defined distance of one another to the same cluster group. For the poplar plantation, trees are approximately 6m apart, therefore a clustering distance of points within 3m was used so as not to introduce too much potential for incorrectly designating trees to neighbouring clusters. A minimum number of 3 observations also had to fall within this proximity for it to be deemed a cluster, to have at least some redundancy and be able to form a polygon to centralize an average point. Tables 4.4, 4.5, and 4.6 present the number of clustered, individual tree positions calculated for each test path of tests 1, 2, and 3 respectively.

Table 4.4 Test 1 consolidated tree positions

Test 1	Conf ≥ 50 Dist ≤ 15 # Trees reduced from clusters	Conf ≥ 70 Dist ≤ 8 # Trees reduced from clusters
1_ A to B	20	3
2_ B to A	8	9
3_ Spiral	62	30
4_ CtoA snake	17	13
5_ BtoC	*GPS data unreliable*	*GPS data unreliable*
6_ BtoC evening	25	10
7_ CtoB	*GPS data unreliable*	*GPS data unreliable*
8_ CtoB evening	19	7
9_ AtoB fast	16	1

Table 4.5 Test 2 consolidated tree positions

Test 2	Conf ≥ 50 Dist ≤ 15 # Trees reduced from clusters	Conf ≥ 70 Dist ≤ 8 # Trees reduced from clusters
11_ A to B	26	14
12_ B to A	18	15
13_ Spiral	55	46
14_ Snake	25	19
15_ A to B Fast	22	9
16_ B to C	31	9
17_ C to B	*GPS data unreliable*	*GPS data unreliable*
18_ A to B no vid	24	17

Table 4.6 Test 3 consolidated tree positions

Test 3	Conf ≥ 50 Dist ≤ 15 # Trees reduced from clusters	Conf ≥ 70 Dist ≤ 8 # Trees reduced from clusters
21_ A to B	8	3
22_ B to A	8	4
23_ A to B two	11	5
24_ B to A two	8	1

Figure 4.11 represents an example of these results for test path A to B during Test 2 with the filtering of 50% confidence and 15m distance range, while Figure 4.12 represents the same test path with filtering of 70% confidence and 8m distance range. Figure 4.13 presents a direct comparison of the consolidated tree positions overlaid on all the individual observations before clustering for the 50% confidence and 15m distance.

Figure 4.11 Test 2 path A to B consolidated tree positions with filtering of 50% confidence and 15m distance range

Figure 4.12 Test 2 path A to B consolidated tree positions with filtering of 70% confidence and 8m distance range

51
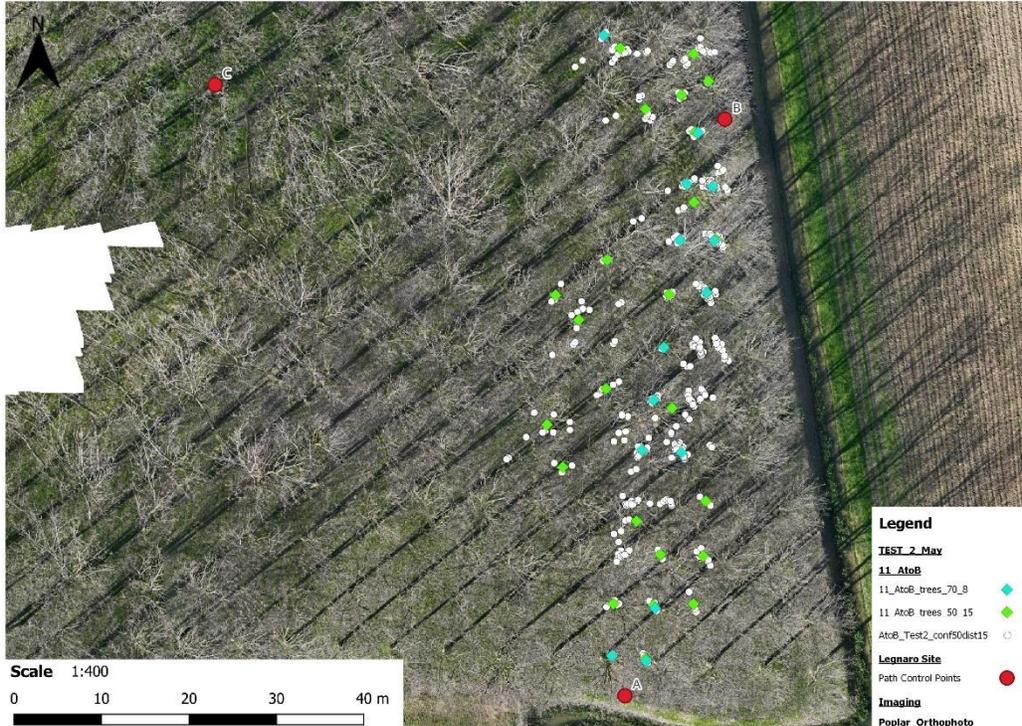
Figure 4.13 Test 2 path A to B consolidated tree positions overlayed on individual tree observation positions with filtering of 50% confidence and 15m distance range

Figure 4.14 visually represents each path's total number of observations during the test run (primary Y axis) and number of consolidated tree positions it reduced to (secondary Y axis) after the refinement both of confidence greater than 50 and distance within 15m, and further refinement of confidence greater than 70 and distance within 8m. Tests that were deemed unreliable are entered with 0 observations.

	A to B	B to A	Spiral	CtoA snake	BtoC	BtoC evening	CtoB	CtoB evening	AtoB fast	A to B	B to A	Spiral	Snake	A to B Fast	B to C	C to B	A to B no vid	A to B	B to A	A to B two	B to A two
# Reliable Obs.	1561	2226	2808	1574	0	1217	0	921	881	3681	4031	14133	5915	3430	5465	0	3733	604	2485	722	1980
Final Trees 50	20	8	62	17	0	25	0	19	16	26	18	55	25	22	31	0	24	8	8	11	8
Final Trees 70	3	9	30	13	0	10	0	7	1	14	15	46	19	9	9	0	17	3	4	5	1
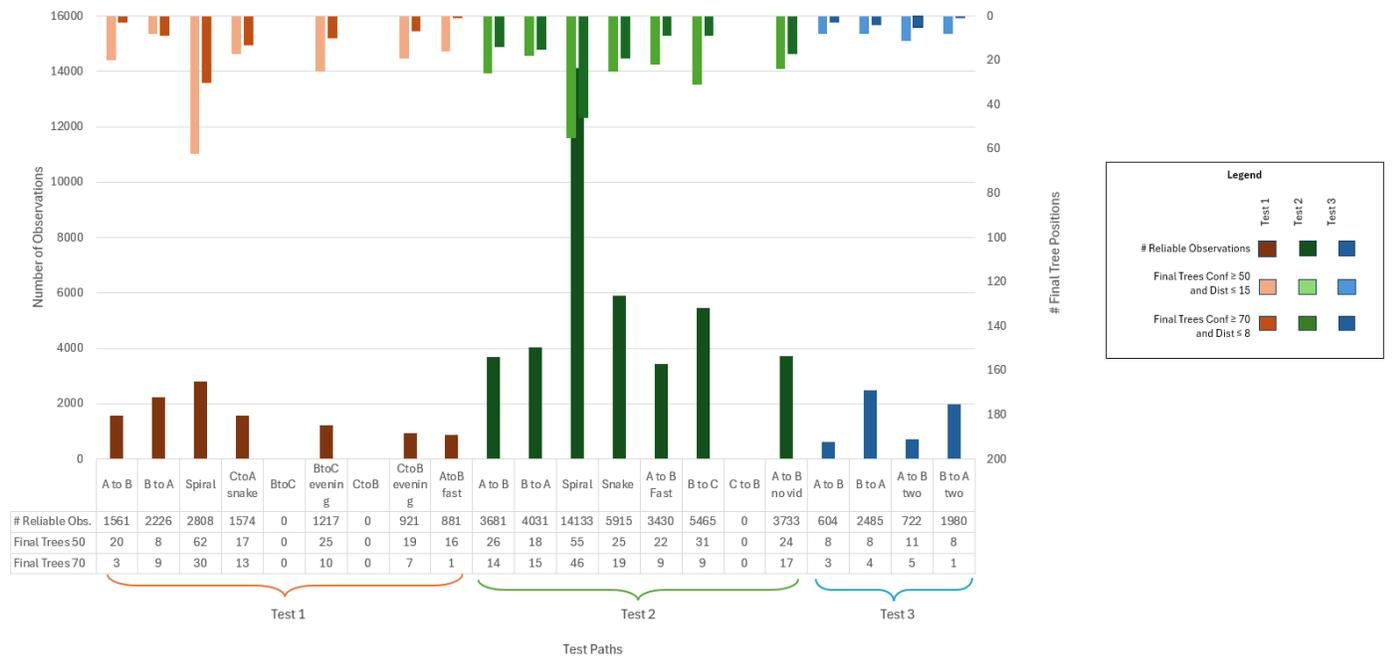
Figure 4.14 Comparison of test runs, number of observations, and number of final tree positions after both stages of refinement

5.0 Discussion

This section will discuss the significance of the results presented in section 4.0, both conceptually and with statistically relevant information determined. It will discuss any limitations encountered or determined, as well as future developments that would be beneficial.

5.1 Test Route Types

As presented in section 3.1.2, three main test route types were introduced. The linear path, anticipated to garner the most clear and accurate results, the snake path, adding more complexity to the movement of the stereo camera through the test site, and the spiral path, with the highest amount of object repetition and directional change. In looking at the comparison of the number of observations taken in total during a test versus the number of observations remaining after refinement of both confidence greater than 50 and distance within 15m, and confidence greater than 70 and distance within 8m, there was an increase in percentage of observations meeting the refinement threshold for snake and spiral path types compared to linear in Test 1. For example, confidence greater than 70 and distance within 8 m, linear paths A to B and B to A saw 1.18% and 5.08% of their observations kept respectively, while snake and spiral paths saw 8.40% and 7.29% of their results kept respectively. Despite anticipating the higher number of initial observations, it was not anticipated that the retention rate of refined observation data would be of higher percentage for the more complex path routes. For Test 2, these results were a lot closer in range with linear paths A to B and B to A seeing 5.02% and 5.76% of their observations kept respectively, while snake and spiral paths seeing 5.57% and 6.48% of their results kept respectively.

5.2 Test Speed

During Test 1 and Test 2, linear path A to B was performed both at the normal pace (0.6 m/s) and at a faster pace (1.2 m/s). (Göncz & Majdik, 2022) referenced that "*The number of objects in the environment, however, directly influenced the time needed for the data association step (15 ms per object).*" And while such reference was to do with the "busyness" of a scene and the time it would need to take in each individual object should there be many in the field of view, it establishes the basic relationship between time and quality/quantity of object observation, which set an expectation for this study that increasing speed of movement through the test site, and reducing the stereo camera's time to take in the entirety of a frame before trying to move on, would directly negatively impact the quantity and quality of observations obtained.

In looking at the comparison of the number of observations taken in total during a test versus the number of observations remaining after refinement of both confidence greater than 50 and distance within 15m, and confidence greater than 70 and distance within 8m, there was a decrease in number of observations in general obtained during the faster test run for both Test 1 and Test 2. Test 1 saw a decrease from 1561 to 881 observations, while Test 2 saw a decrease from 3681 to 3430 observations. However, the percentage of observations meeting the refinement threshold both for the broader refinement and more reduced refinement actually increased rather than decreased with speed. For the broader refinement of confidence greater than 50 and distance within 15m, Test 1 saw a retention increase from 14.59% to 15.73%, and Test 2 saw a retention increase from 1.18% to 1.30%. For the more reduced refinement of confidence greater than 70 and distance within 8m, Test 1 saw a retention increase from 28.23% to 35.57%, and Test 2 saw a retention increase from 5.02% to 6.05%.

5.3 Accuracy of Calculated Tree Positions

From the Legnaro test site, accuracy of the calculated consolidated tree positions can be analysed by comparing those calculated tree positions to the tree positions extracted from the drone orthophoto. To test the proximity of these points, buffers were generated around the "accurate" tree positions drone orthophoto ranging from 1m to 3m in radius (Fig. 5.1).

Test path A to B during Test 2 was used for this accuracy test, and results of the amount of calculated tree positions falling within each of the buffer ranges is presented in Table 5.1.

Table 5.1 Calculated tree positions within buffer ranges of 1 to 3m

	Conf ≥ 50 Dist ≤ 15		Conf ≥ 70 Dist ≤ 8	
	# Trees Calced	26	# Trees Calced	14
Buffer Size:	Trees within buffer:	% of trees in buffer	Trees within buffer:	% of trees in buffer
1m	5	19%	7	50%
1.5m	10	38%	12	86%
2m	13	50%	14	100%
2.5m	15	58%	14	100%
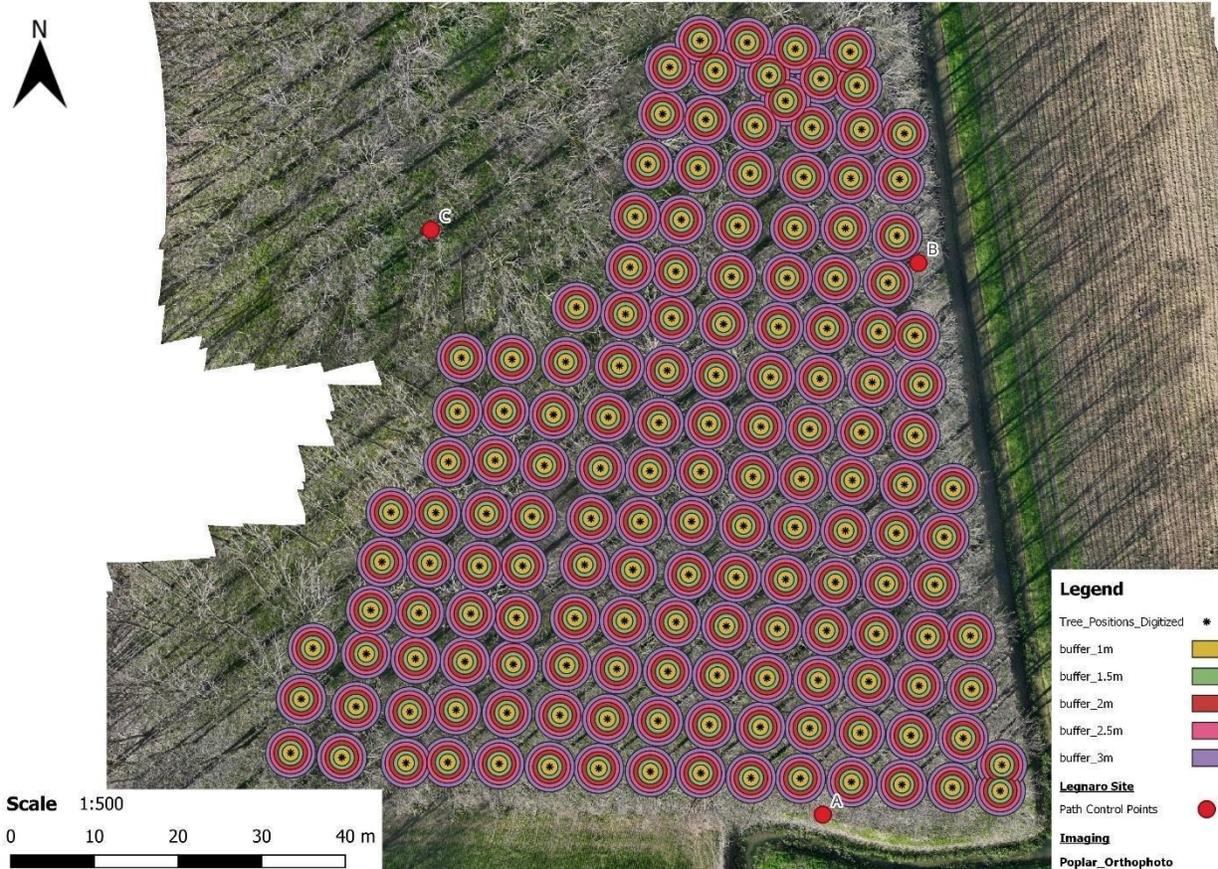
Figure 5.1 Buffers of 1 to 3m around extracted tree positions from zone data

Figures 5.2 and 5.3 illustrate the calculated tree positions overlayed on a 2m buffer, which saw a 50% "accuracy" rate for the broader refinement of confidence greater than 50 and distance within 15m, and 100% accuracy for the more reduced refinement of confidence greater than 70 and distance within 8m. This provides indication that with appropriate refinement of observation data, at this point in stereo camera training and testing calculation procedures, tree positions can be expected to be relatively accurate within 2m. Seeing as though diameters of trees were ranging between 0.3-0.5m and the positive control tree positions were extracted from a singular drone flight's orthomosaic, this provides a good foundation for data processing.
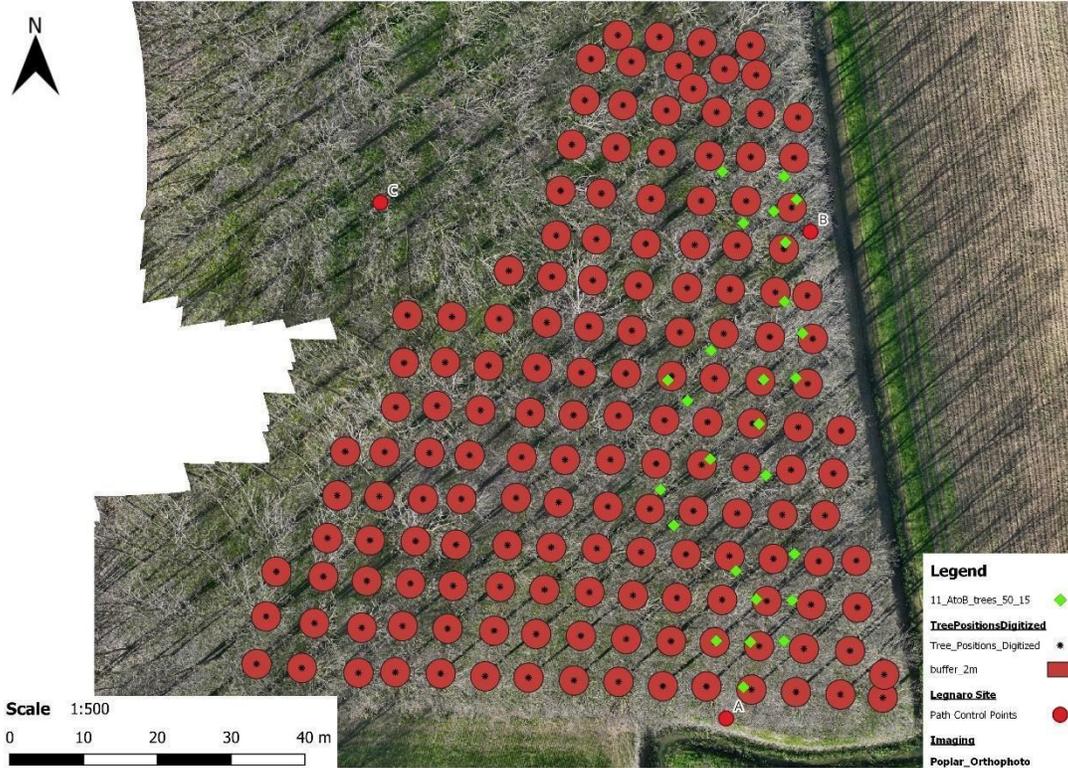
Figure 5.2 Overlay of calculated tree positions for broader refinement on 2m buffer
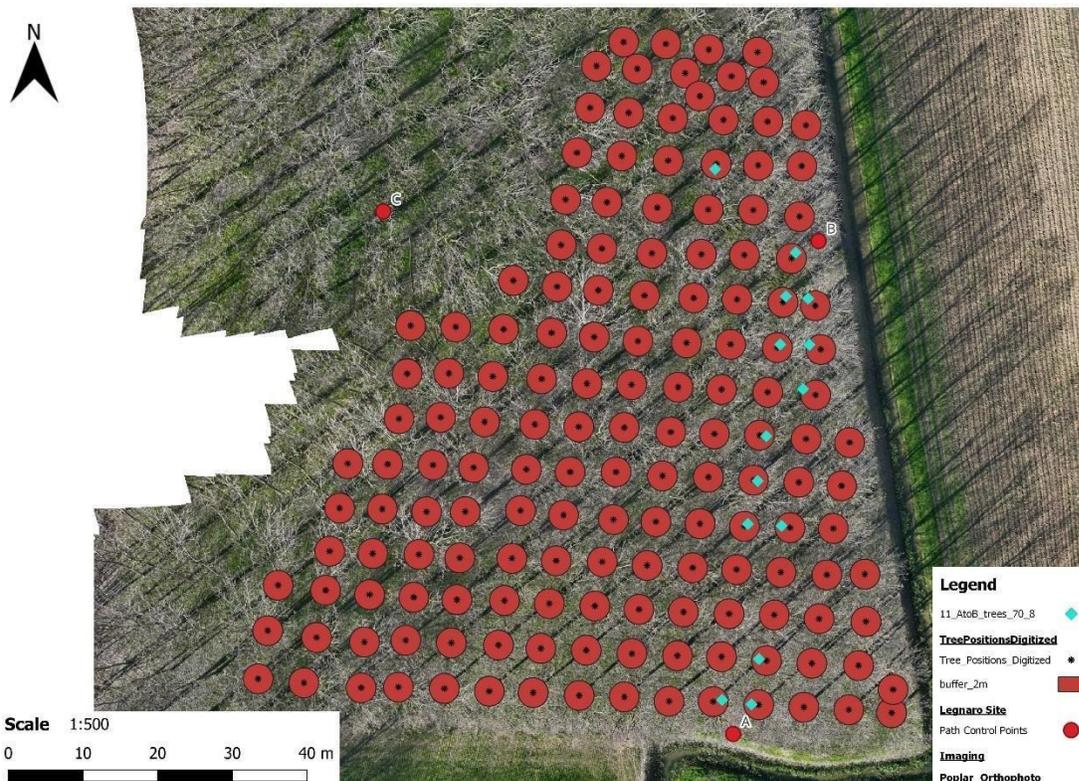
Figure 5.3 Overlay of calculated tree positions for reduced refinement on 2m buffer

5.4 Role of Lighting and Shadows

Test 2, performed in May once trees had foliage, saw an overall increase in observations made by the ZED 2i stereo camera for each test path run compared to the same test routes performed in the same trees during Test 1. Figure 5.4 represents comparison of only test paths that were repeated during Test two and both tests provided reliable results. Test 1 was performed in February when there were not yet leaves on the trees, indicating that lighting can play a role on the object detection capabilities of the stereo camera. This conclusion is concurrent with that made in Liszka et al. (2023), that saw depth maps containing less NaN values in cloudy images taken vs. images taken when it was sunny, indicating that objects were easier to define when the sky was cloudier.
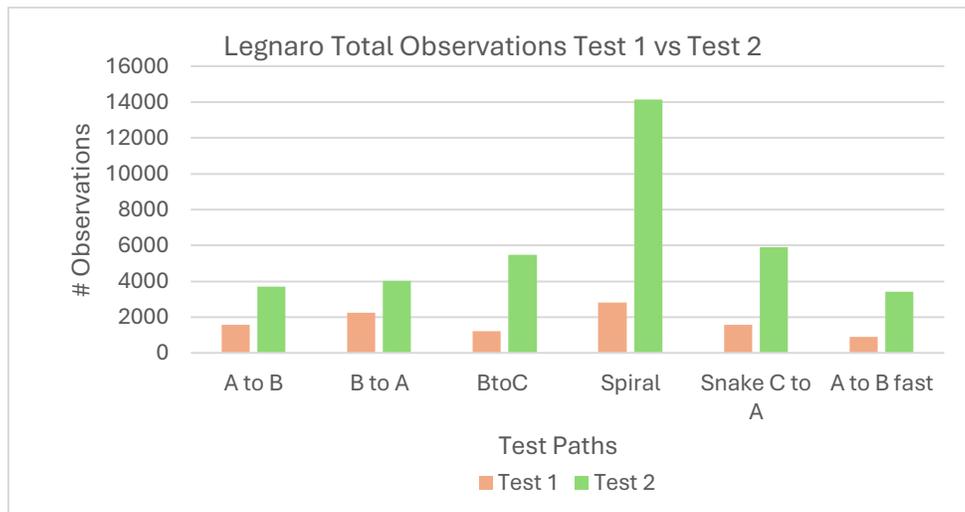
Figure 5.4 Comparison of number of total observation positions Test 1 vs Test 2

Figure 5.5 represents comparison of the final number of tree positions calculated during these two tests, demonstrating that save the spiral path, all other paths that had more initial observations did present more final tree positions. It is, however, not a direct assumption that can be made that more observations lead to more refined tree positions, as an increase in general observations can just be repetitive observations of the same trees previously detected. As the spiral path experiences the most repetition in detecting same trees, this could be a factor that explains why Test 2 saw a 7 tree decrease in final free positions consolidated from data, despite having a general number of detections 5x greater than Test 1. This highlights that the number of detections made during a test is not a direct reflection of the quality of the data, nor does it represent how

many of those observations are redundant and creating larger clusters for the same consolidated final trees as opposed to different tree positions.
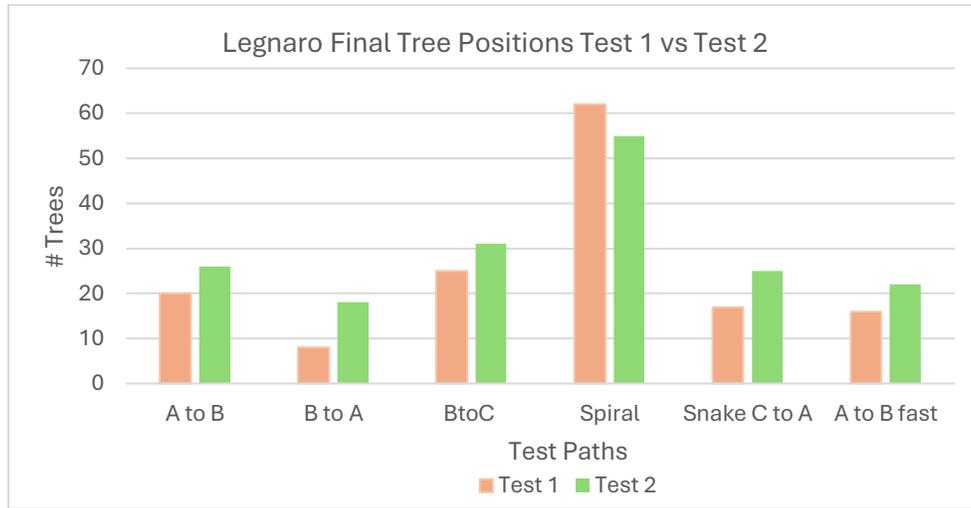
Figure 5.5 Comparison of total number of final tee positions calculated Test 1 vs. Test 2

5.5 Organized vs. Natural Forest Environment

The natural forest environment used a singular test path and performed four tests. This includes two test paths going from Control Point A to B, and two going from B back to A. Notably, the tests from B to A saw an increase of 2-3x the number of observations as the tests from A to B (Fig. 5.6).
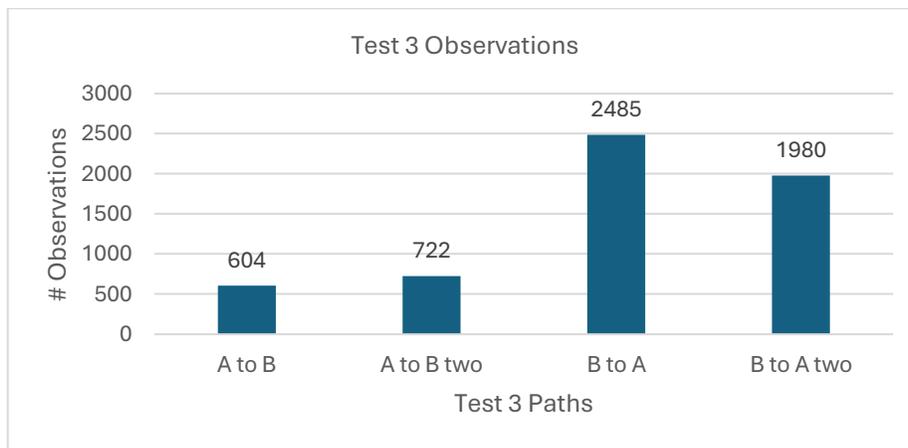
Figure 5.6 Test 3 comparison of observations made during A to B path vs B to A path

In looking at the final number of consolidated trees, again there is no direct correlation between number of observations and number final tree positions deduced, as final results for tree positions, particularly when looking at the broader refinement of observations with confidence

greater than 50 and distance within 15m are consistent between paths going A to B vs. B to A (Fig. 5.7).
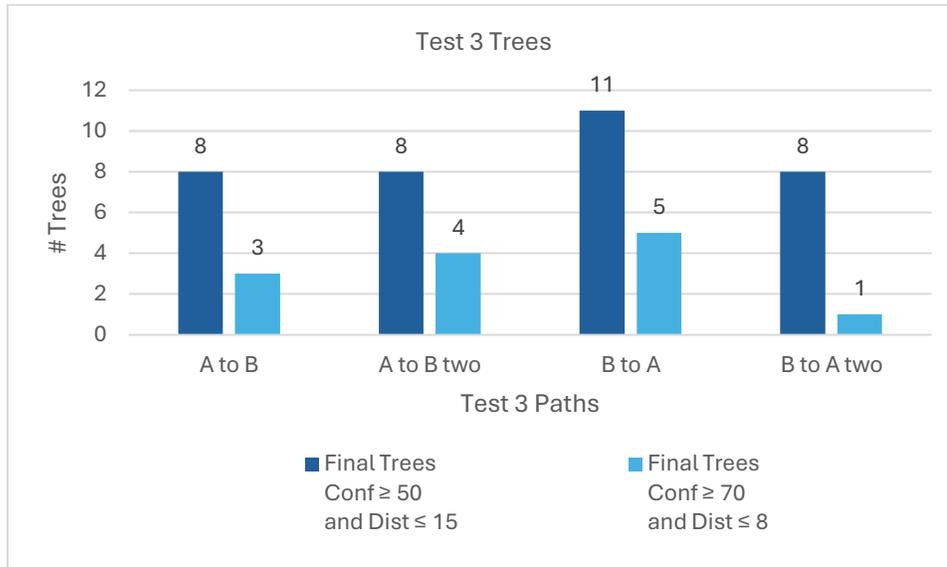
Figure 5.7 Test 3 final number of tree positions after observation refinement and consolidation

To compare this Taibon test site from Test 3 to the Legnaro test site from Tests 1 and 2, a comparison of refined observations will be highlighted (Fig. 5.8). At the highest level of refinement for individual detections of confidence greater than 70 and distance within 8m, only the generally linear paths will be compared. This figure directly compares the total number of observations per test path during the different tests to the number of observations that remained from those paths after refinements. On the secondary axis, this figure summarizes the average percentage of surviving observations per test, calculated by averaging the percentage of remaining observations of each test path with percentages of all paths in consideration from the same test. Test 1 and 2 in Legnaro are represented by orange and green respectively, with Test 3 in Taibon being blue. Test 1 had an overall average retention rate of 4.43% of original observations, while Test 2 and Test 3 had a retention rate of 5.56 and 4.20% respectively.

This corresponds with the conclusions of section 5.4, that Test 2 performed in May once there was tree foliage yielded more results than Test 1 in February when there was yet to be any tree foliage. It also highlights Test 3 as having the lowest retention rate of observations after refinement took into consideration the detection characteristics influencing potential accuracy (confidence and distance). This presents that a more natural, busy forest environment may see less

observations making it through the refinement process. As highlighted by Göncz & Majdik (2022) "*a crowded environment with many objects in close proximity to each other is a limiting factor*."
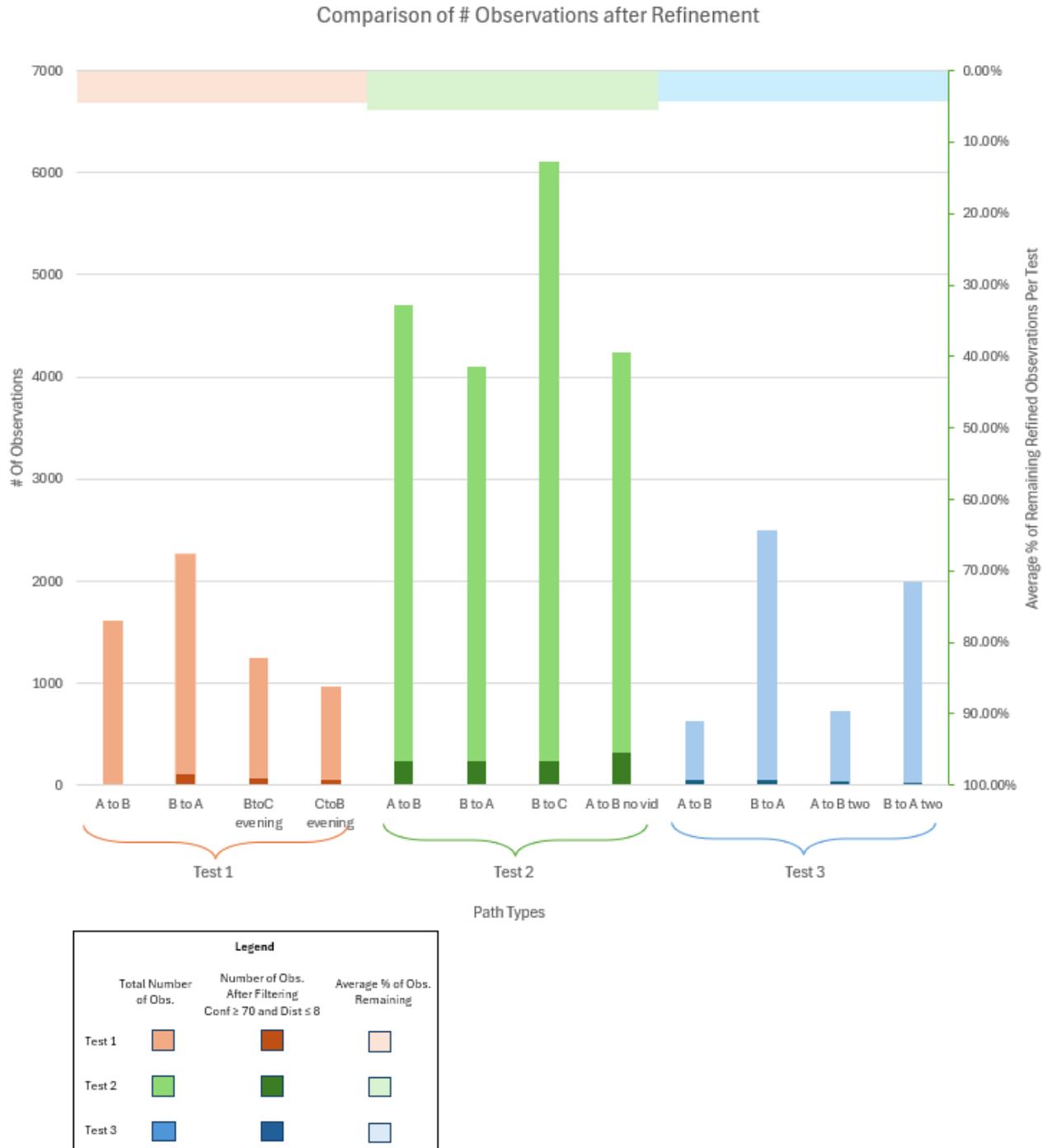
Figure 5.8 Comparison of retention of observations after refinement processes

5.6 Limitations of Work

This section will contain a breakdown of assumptions or limitations encountered throughout the work of this thesis, highlighting room for improvement as research on the subject progresses.

An element not taken into consideration in comparing accuracies between Test 1 and 2 was the difference that may have been influenced by the difference in stereo camera lens due to availability of equipment during the time of the surveys. As mentioned in section 2.2.1, Test 1 and 3 used a ZED 2i stereo camera with a 2.1mm lens, while Test 2 used a 4mm lens. For the purpose of this work, that was assumed to be negligible.

Another limitation not thoroughly investigated was the limitation presented by working with GPS in forested areas. This setting made it difficult to maintain fixed positioning, often resulting in float or single observations which resulted in some unreliable test paths being excluded due to insufficient accuracy. Brief assessments were made on the composition of the solution of points along the paths, but no thorough assessment of reliability or accuracy was made beyond looking at the paths to see if they seemed reasonable. As global positions were calculated based on this positional data, taking accuracy of these points into further consideration could affect the final accuracy of calculated tree positions.

With respect to the calculation of tree positions, IMU data was not taken into consideration. This preliminary calculation procedure made the assumption that the direction of the camera could be determined by averaging the next five positions logged in front of a given point, and that the camera was exactly straight and level during the entirety of the test.
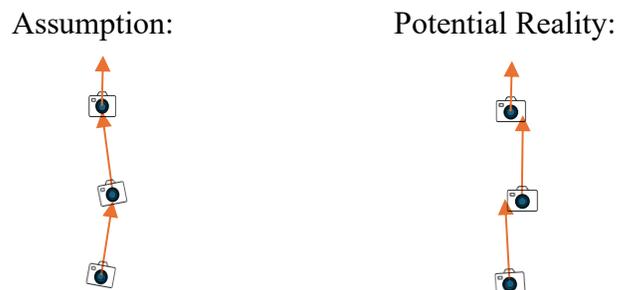
Figure 5.9 Camera positioning assumption

Acknowledging and using IMU data would more properly take into account the orientation and potential rotation of the camera throughout the test runs in order to more accurately calculate the position of trees observed.

Additional next steps for immediate improvements to this methodology go back to the training stages of the stereo camera as well. Improving the initial training of the stereo camera will improve observation detection. Continued studies will also incorporate more than just trees as identifiable objects. There are many other "objects" present in forest environments, therefore once training and processing procedures are further refined, commencing the training of the ZED 2i Stereo Camera to recognize other objects such as boulders, shrubs, etc. will be greatly beneficial. Further initiatives would also involve incorporating the dimensioning capabilities, in which the tree diameter data collected during the first field survey would come into play.

Finally, a goal would be to upgrade to real-time processing, as that would be a step necessary for the proposed interest of implementation for object detection and obstacle avoidance in forest operations. Taking into consideration these limitations and improvements moving forward will greatly progress this research further.

6.0 Conclusions

This thesis has seen the objectives met of obtaining, training, and establishing a preliminary set of processing procedures for data collected by the ZED 2i Stereo Camera.

The primary objective of this thesis was to progress from the initial stages of training and testing a stereo camera, and to develop a method for converting object observations into precise object position points. Detection software was trained to implement with the ZED 2i stereo camera for data collection and three field survey tests were performed. The first two field tests were conducted in a plantation test area in Legnaro, PD without and then with tree foliage present. This change in environmental condition saw the increase in object detections made by the stereo camera during all test paths. A third test was then conducted in a natural forest test area in the Taibon Agordino region of Belluno. This change in test location saw a reduction in retention of observations going through the same refinement procedures as Test 1 and 2 based on detection confidence and distance from the stereo camera to the observation position.

This data was used to configure processing procedures to allow for the extraction of meaningful results, which allowed for an opening assessment of the performance of the processing methods established against control data. Observation data in a local coordinate system from a given test run can be matched with GNSS global positions of the camera throughout the test run, allowing for the calculation of the global coordinates of any given observation. These clusters of observations that include many repeated detections of same trees can then be refined based on accuracy, and consolidated into a single tree position per group of observation clustering.

Although there is considerable room for enhancing the efficiency of the calculations and incorporating additional data (such as GPS positional accuracies at the time of observation and IMU data for accurate camera positioning), this work represents a foundational effort in refining raw data into interpretable results. This initial attempt lays the groundwork for future improvements in data interpretation and accuracy.

References

Anne Almendral, K. M., Mae Babaran, R. G., Jones Carzon, B. C., Patrick Cu, K. K., Lalanto, J. M., & Abad, A. C. (2018). Autonomous Fruit Harvester with Machine Vision. *Journal of Telecommunication, Electronic and Computer Engineering*, *10*(1–6). www.stereolabs.com

Arunpriyan, J., Variyar, V. V. S., Soman, K. P., & Adarsh, S. (2020). Real-Time Speed Bump Detection Using Image Segmentation for Autonomous Vehicles. *Advances in Intelligent Systems and Computing*, *1039*, 308–315. https://doi.org/10.1007/978-3-030-30465-2_35

Bin Mohammed Na'aim, M. Z., & Abdul Manaf, M. B. (2024). Establishment of control points using GNSS-RTK technique. *E3S Web of Conferences*, *479*. https://doi.org/10.1051/e3sconf/202447902001

Chris Veness. (2022). *Calculate distance, bearing and more between Latitude/Longitude points*. https://www.movable-type.co.uk/scripts/latlong.html

da Silva, D. Q., dos Santos, F. N., Filipe, V., Sousa, A. J., & Oliveira, P. M. (2022). Edge AI-Based Tree Trunk Detection for Forestry Monitoring Robotics. *Robotics*, *11*(6). https://doi.org/10.3390/robotics11060136

Dendrotik. (2016). *TREE CALIPER*. https://www.dendrotik.com/en/product/bahco-tree-caliper/#

Emlid. (2019). *REACH M2*. https://emlid.com/reach/

Emlid. (2022). *REACH RS2+*. https://emlid.com/reachrs2plus/

Fleischmann, P., & Berns, K. (n.d.). *A Stereo Vision Based Obstacle Detection System for Agricultural Applications*.

Göncz, L., & Majdik, A. L. (2022). Object-Based Change Detection Algorithm with a Spatial AI Stereo Camera. *Sensors*, *22*(17). https://doi.org/10.3390/s22176342

Haglof Sweden. (n.d.). *Vertex Laser Geo*. Retrieved June 20, 2024, from https://haglofsweden.com/project/laser-geo/

Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision*.

Jrondi, Z., Moussaid, A., & Hadi, M. Y. (2024). Exploring End-to-End object detection with transformers versus YOLOv8 for enhanced citrus fruit detection within trees. *Systems and Soft Computing*, *6*. https://doi.org/10.1016/j.sasc.2024.200103

Kao, S. M., Ma, C. W., & Hsieh, J. W. (2022). Integrating Low-Cost LiDAR and Stereo Camera for 3D Object Detection with Lidar-Stereo Fusion and Cost Volume Enhancement. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, *2022-November*. https://doi.org/10.1109/TENCON55691.2022.9977732

Lagisetty R., Philip N. K., Padhi R., & Bhat M. S. (2013). *Object Detection and Obstacle Avoidance for Mobile Robot using Stereo Camera*.

Li, S., & Lideskog, H. (n.d.). *FOREST TERRAIN OBJECT DETECTION BASED ON RGB AND DEPTH INFORMATION*.

Liszka, S., Södergren, H., Löfstedt, T., & Björklund, H. (2023). *Evaluation of Tree Planting using Computer Vision models YOLO and U-Net*.

McGlade, J., Wallace, L., Hally, B., Reinke, K., & Jones, S. (2023). The Effect of Surrounding Vegetation on Basal Stem Measurements Acquired Using Low-Cost Depth Sensors in Urban and Native Forest Environments. *Sensors*, *23*(8). https://doi.org/10.3390/s23083933

MedCalc Software Ltd. (2024). *ATAN2 function*. https://www.medcalc.org/manual/atan2-function.php

Omni Calculator. (2024). *Azimuth Calculator*. https://www.omnicalculator.com/other/azimuth

OpenCV. (2024). *OpenCV*. https://opencv.org/

PyTorch. (2024). *PyTorch*. https://pytorch.org/

QGIS. (2024). https://www.qgis.org/en/site/

roboflow. (2023a). *Tree Computer Vision Project*. https://universe.roboflow.com/12345678-i0yth/tree-relit

roboflow. (2023b). *tree detection model Computer Vision Project*.

R-Project. (2024). *The R Project for Statistical Computing*. https://www.r-project.org/

STEREOLABS. (2022). *ZED 2i*. www.stereolabs.com

Stereolabs Inc. (2024a). *3D Object Detection Overview*. https://www.stereolabs.com/docs/object-detection

Stereolabs Inc. (2024b). *Depth Sensing Overview*. https://www.stereolabs.com/docs/depth-sensing

Stereolabs Inc. (2024c). *Recommended Specifications for ZED SDK*. https://www.stereolabs.com/docs/installation/specifications

Stereolabs Inc. (2024d). *Setting up GNSS / RTK*. https://www.stereolabs.com/docs/global-localization/using-gnss-linux#enabling-rtk-on-your-gnss-module

Wells, L. A., & Chung, W. (2023). Vision-Aided Localization and Mapping in Forested Environments Using Stereo Images. *Sensors*, *23*(16). https://doi.org/10.3390/s23167043

Annexes

This section includes a summary of abbreviations introduced throughout the text, as well as the tree position calculations introduced (Eq. 3.1-3.6) converted into Microsoft Excel commands.

Summary of Acronyms and Abbreviations

The following is a list of the main acronyms and abbreviations used throughout this paper, in order of appearance.

GNSS: Global Navigation Satellite System

LiDAR: Light and Detection Ranging

TLS: Terrestrial Laser Scanning

RGB-D: Red Green Blue, Depth (sensor)

IMU: Inertial Measurement Unit

SDK: Software Development Kit

YOLO: You Only Look Once

PSF: Python Software Foundation

GIS: Geographic Information Systems

DBH: Diameter At Breast Height

OpenCV: Open-Source Computer Vision Library

RTK: Real-Time Kinematic

Excel Conversion Calculation Spreadsheet: Column Descriptions and Formulas:

Column A: Adjusted Time

- Adjusted timestamp for the RTK GNSS data, starting at start time of stereo camera data collection.
  = *imported*

Column B: Lat

- Latitude of each RTK GNSS point taken and within the timeframe of the stereo camera survey.
  = *imported*

Column C: Long

- Longitude of each RTK GNSS point taken and within the timeframe of the stereo camera survey.
  = *imported*

Column D: Ellipsoidal Height

- Ellipsoidal height of each RTK GNSS point taken and within the timeframe of the stereo camera survey.
  = *imported*

Column E: Object ID

- Object ID # of every single object identified.
  = *imported*

Column F: Confidence

- Confidence of every single object identified, 0-100, 100 being most accurate.
  = *imported*

Column G: Distance

- Distance of every single object identified from standing position of stereo camera.
  = *imported*

Column H: ΔX

- Change on the local X axis of every single object identified from standing position of stereo camera.
  = *imported*

Column I: ΔY

- Change on the local Y axis of every single object identified from standing position of stereo camera.
  = *imported*

```
Column J: ΔZ
```

- Change on the local Z axis of every single object identified from standing position of stereo camera.
  = *imported*

Column K: Timestamp (+18.6 sec)

- Adjusted timestamp for each observation that was taken.
  = *imported*

Column L: Row # of Matching GNSS Data Timestamp

- Determining the row of the matching RTK GNSS data for each observation timestamp.
  =ROW(INDEX($C$2:$C$2264, MATCH([@[Timestamp (+18.6 sec]],$A$2:$A$2264, 0)))

Column M: Lat. at ID Observations Point

- Collecting the latitude of the matching RTK GNSS point for each observation timestamp to use as zero position of stereo camera.
- *=INDEX(B:B, [@[Row '# of Matching GNSS Data Timestamp]])*

Column N: Long. at ID Observations Point

- Collecting the longitude of the matching RTK GNSS point for each observation timestamp to use as zero position of stereo camera.
- *=INDEX(C:C, [@[Row '# of Matching GNSS Data Timestamp]])*

Column O: Lat next GNSS Point

- Collecting the latitude of the next five points *after* the matching RTK GNSS point and averaging them for each observation timestamp to use to calculate assumed general direction of movement and therefore assumed general direction stereo camera is pointing.
- *=AVERAGE(INDEX($B$1:$B$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+1)), INDEX($B$1:$B$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+2)), INDEX($B$1:$B$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+3)), INDEX($B$1:$B$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+4)), INDEX($B$1:$B$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+5)))*

Column P: Long next GNSS Point

- Collecting the longitude of the next five points *after* the matching RTK GNSS point and averaging them for each observation timestamp to use to calculate assumed general direction of movement and therefore assumed general direction stereo camera is pointing.
- *=AVERAGE(INDEX($C$1:$C$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+1)), INDEX($C$1:$C$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+2)), INDEX($C$1:$C$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+3)), INDEX($C$1:$C$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+4)), INDEX($C$1:$C$1612, ([@[Row '# of Matching GNSS Data Timestamp]]+5)))*

Column Q: Δ Lat

- Calculating the change between two consecutive latitude points, intermediate step to calculate assumed direction of movement and therefore assumed direction stereo camera is pointing.
- *=[@[Lat next GNSS Point]]-[@[Lat. at ID Observations Point]]*

Column R: ΔLong

- Calculating the change between two consecutive longitude points, intermediate step to calculate assumed direction of movement and therefore assumed direction stereo camera is pointing.
- *=[@[Long next GNSS Point]]-[@[Long. at ID Observations Point]]*

Column S: Atan2 X component (rad)

- Calculating the x component for ATAN2 calc in order to calculate azimuth of assumed direction of movement and therefore assumed direction stereo camera is pointing.
- *=(SIN(RADIANS(R2))*COS(RADIANS(O2)))*

Column T: Atan2 Y component (rad)

- Calculating the x component for ATAN2 calc in order to calculate azimuth of assumed direction of movement and therefore assumed direction stereo camera is pointing.
- *=(COS(RADIANS(M2))*SIN(RADIANS(O2))-SIN(RADIANS(M2))*COS(RADIANS(O2))*COS(RADIANS(R2)))*

Column U: Arctan2 corrected

- Calculating the ARCTAN2 angle of assumed direction of movement and therefore assumed direction stereo camera is pointing. Formula conditional to direction of movement.
- *=IF((DEGREES(ATAN2([@[Atan2 X component (rad)]],[@[Atan2 Y component (rad)]]))<0), DEGREES(ATAN2([@[Atan2 X component (rad)]],[@[Atan2 Y component (rad)]]))+360, DEGREES(ATAN2([@[Atan2 X component (rad)]],[@[Atan2 Y component (rad)]])))*

Column V: Azimuth (True)

- Calculating the actual azimuth angle of assumed direction of movement and therefore assumed direction stereo camera is pointing. Formula conditional to direction of movement.
- *=IF((([@[Arctan2 corrected]]<90),90-[@[Arctan2 corrected]], 450-[@[Arctan2 corrected]])*

Column W: Rotation (β)

- Calculating the rotation angle of the local coordinate system to nearest previous axis to local North (-z). Formula conditional to direction of movement.
- *=IF(V2>270, V2-270, IF(V2>180, V2-180, IF(V2>90, V2-90, V2)))*

Column X: ΔΦ lat

- Calculating the change in latitude between observed object and position at time of observation based on rotation angle β and local coordinate system position changes ΔX, ΔY, ΔZ.

- *=IF([@[Azimuth (True)]]>270, ((-[@ΔZ])*SIN(RADIANS([@[Rotation (β)]])))+([@ΔX]*COS(RADIANS([@[Rotation (β)]]))), IF([@[Azimuth (True)]]>180, -(((-[@ΔZ])*COS(RADIANS([@[Rotation (β)]])))-([@ΔX]*SIN(RADIANS([@[Rotation (β)]])))), IF([@[Azimuth (True)]]>90, -(((-[@ΔZ])*SIN(RADIANS([@[Rotation (β)]])))+([@ΔX]*COS(RADIANS([@[Rotation (β)]])))), ((-[@ΔZ])*COS(RADIANS([@[Rotation (β)]])))-([@ΔX]*SIN(RADIANS([@[Rotation (β)]]))))))*

Column Y: Δλ long.

- Calculating the change in longitude between observed object and position at time of observation based on rotation angle β and local coordinate system position changes ΔX, ΔY, ΔZ.

- *=IF([@[Azimuth (True)]]>270, -((((-[@ΔZ])*COS(RADIANS([@[Rotation (β)]])))-([@ΔX]*SIN(RADIANS([@[Rotation (β)]]))))), IF([@[Azimuth (True)]]>180, -(((-[@ΔZ])*SIN(RADIANS([@[Rotation (β)]])))+([@ΔX]*COS(RADIANS([@[Rotation (β)]])))), IF([@[Azimuth (True)]]>90, (((-[@ΔZ])*COS(RADIANS([@[Rotation (β)]])))-([@ΔX]*SIN(RADIANS([@[Rotation (β)]])))), ((-[@ΔZ])*SIN(RADIANS([@[Rotation (β)]])))+([@ΔX]*COS(RADIANS([@[Rotation (β)]]))))))*

Column Z: ID Lat.

- Adding the change in latitude to the latitude position of RTK GNSS observation at zero point from which an observation was taken to get the global longitudinal position of the observed object.

- =[@[Lat. at ID Observations Point]]+(([@[ΔΦ lat]]/1000)/(6378))*(180/PI())

Column AA: ID Long.

- Adding the change in longitude to the longitude position of RTK GNSS observation at zero point from which an observation was taken to get the global longitudinal position of the observed object.

- =[@[Long. at ID Observations Point]]+(([@[Δλ long.]]/1000)/(6378))*(180/PI())/COS(RADIANS([@[Lat. at ID Observations Point]]*(PI()/180)))