



**Budapest University of Technology and Economics**

Faculty of Electrical Engineering and Informatics

Department of Measurement and Information Systems

# **Drone Localization in Ad-hoc Indoor Environment**

MASTER'S THESIS

*Author*

Marcell Rausch

*Advisor*

dr.Gábor Fehér

May 28, 2020

# Contents

<b>Kivonat</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Literature review</b>	<b>4</b>
2.1 Basic aviation terminologies . . . . .	4
2.1.1 UAV or Drone . . . . .	4
2.1.2 Body axes used in aviation . . . . .	4
2.1.3 Ground Control Station . . . . .	5
2.1.4 Multicopters . . . . .	5
2.2 Simultaneous Localization And Mapping - SLAM . . . . .	5
2.2.1 Visual SLAM . . . . .	7
2.2.1.1 Monocular camera . . . . .	8
2.2.1.2 Stereo camera . . . . .	8
2.2.1.3 LSD-SLAM . . . . .	9
2.2.1.4 ORB-SLAM . . . . .	10
2.2.2 Ranging sensor based SLAM . . . . .	11
2.2.2.1 3D LIDAR SLAM to improve localization accuracy . . . . .	11
2.2.2.2 3D LIDAR SLAM for mapping . . . . .	12
2.2.3 Cartographer SLAM . . . . .	13
2.3 Similar products available on the market . . . . .	15
2.3.1 Terabee TeraRanger Tower . . . . .	15
2.3.2 Crazyflie Multi-ranger deck . . . . .	16
2.3.3 Skydio 2 . . . . .	17
2.3.4 Velodyne Velabit . . . . .	17
2.4 Components review . . . . .	18
2.4.1 Robot Operating System . . . . .	18

2.4.2	Gazebo Simulator . . . . .	19
2.4.3	PX4 Autopilot Software . . . . .	20
2.4.3.1	PX4 simulation using MAVROS . . . . .	20
2.4.3.2	PX4 simulation using Gazebo with ROS wrapper . . . . .	20
2.4.4	QGroundControl . . . . .	22
2.5	VL53L1X LIDAR sensor . . . . .	22
<b>3</b>	<b>System design</b>	<b>24</b>
3.1	VL53L1X measurements . . . . .	24
3.1.1	Selected operating modes . . . . .	25
3.2	LIDAR layout . . . . .	27
3.2.1	Layout design for 2D SLAM . . . . .	27
3.2.2	Layout design for 3D SLAM . . . . .	28
3.3	Data collection in Gazebo Simulator . . . . .	28
3.3.1	Development environment in Visual Studio Code . . . . .	29
3.3.2	Adding LIDAR sensor to a quadcopter model . . . . .	30
3.3.3	Building model . . . . .	30
3.3.4	Remote controller . . . . .	31
3.4	Rosbag filtering . . . . .	31
3.4.1	Update rate . . . . .	32
3.4.2	Field of View . . . . .	33
3.4.3	Distance filter . . . . .	34
3.5	Cartographer environment . . . . .	35
3.5.1	Configuration and tuning . . . . .	37
3.5.2	Trajectory extraction . . . . .	38
3.6	SLAM performance evaluation . . . . .	38
3.6.1	Down-sampling . . . . .	39
3.6.2	Offset compensation . . . . .	39
3.6.3	Trajectory error calculation . . . . .	40
3.7	Complete workflow . . . . .	40
3.8	Plans for hardware implementation . . . . .	41
3.8.1	Standalone data collector design . . . . .	41
<b>4</b>	<b>Implementation</b>	<b>43</b>
4.1	VL53L1X measurement setup . . . . .	43
4.1.1	Comparison of selected operating modes . . . . .	44
4.1.2	Detailed distance measurements . . . . .	46

4.1.3	Detailed measurements with the fastest setup . . . . .	48
4.2	2D SLAM evaluation . . . . .	50
4.2.1	Effects of LIDAR resolution . . . . .	51
4.2.1.1	Unfiltered data . . . . .	51
4.2.1.2	2D SLAM with 4x4 resolution . . . . .	52
4.2.1.3	2D SLAM with 3x3 resolution . . . . .	55
4.2.1.4	2D SLAM with 2x2 resolution . . . . .	57
4.2.1.5	2D SLAM with 1x1 resolution . . . . .	58
4.2.1.6	Summary . . . . .	59
4.2.2	Effects of sampling rate . . . . .	59
4.2.2.1	4x4 resolution . . . . .	59
4.2.2.2	3x3 resolution . . . . .	60
4.2.3	Effects of maximum distance . . . . .	61
4.2.4	Reduced set of evenly distributed LIDAR sensors . . . . .	63
4.3	3D SLAM evaluation . . . . .	65
4.3.1	Unfiltered data . . . . .	65
4.3.2	Effects of LIDAR resolution . . . . .	67
4.3.3	Effects of sampling rate . . . . .	69
4.3.4	Effects of maximum distance . . . . .	70
4.3.5	Summary . . . . .	71
4.4	SLAM performance on different building . . . . .	72
<b>5</b>	<b>Conclusion</b>	<b>75</b>

# **Kivonat**

Dolgozatomban bemutatom

# Abstract

In this thesis I propose a system that uses single point LIDAR sensors to scan a multi-copter's environment and build a 3D map of its surroundings using SLAM algorithm. The selected sensor is a VL53L1X time of flight LIDAR sensor developed by STMicroelectronics. Due to recent COVID-19 pandemic, I had no access to the laboratory equipment. As a workaround I have used Gazebo simulator and simulated LIDAR measurements as close real VL53L1X measurements as possible.

Different camera and ranging sensor based solutions can be found for SLAM applications. Ranging sensor Based setups mostly use heavy, sensitive but nonetheless highly accurate planar scanners. A solution for hobbyists and students, called Crazyflie Multi-ranger deck uses the 5 pieces of the same VL53L1X sensor for SLAM applications. Another product called Skydio2 is an autonomous drone that is capable of active tracking and object avoidance even in dense forests, using 6 fisheye cameras to scan its environment. This drone is the most reliable drone for autonomous tracking currently available.

I have used PX4 repository that uses Gazebo to simulate an Iris drone and placed two LIDAR sensors on it, one on the top and one on the bottom. The two sensors make ranges of the surroundings evenly in a sphere shape. LIDAR and IMU measurements are recorded into a Rosbag file. Using a self-developed filtering script, I filter the recorded LIDAR measurements to simulate a number of VL53L1X sensors in predefined orientation and settings. The settings are the resolution, update rate and maximum ranging distance. These parameters come from measurements done with the actual VL53L1X sensor. Offline processing using the Rosbag filter allows to rapidly test sensor settings and different layouts. Because data of the same flight is used for all tests, their performance can be compared.

# Chapter 1

## Introduction

A prerequisite for future expansion of mobile robots is the ability to navigate autonomously and safely in their ever-changing environment without collision. The positioning of drones in ad hoc indoor environment where no infrastructure is available is needed to fulfill this prerequisite. To know the position of a robot in an unknown environment, the robot needs to build a map of its surroundings so its position can be referenced. Mapping and positioning were considered as two separate problems at first and research focused on either of them, but the solution to these problems separately proved to be divergent. Only by considering these two problems together and finding solutions simultaneously, it became convergent [? ]. This process is referred to as Simultaneous Localization And Mapping (SLAM).

The proposed issue by SLAM has many solutions depending on the use-case. SLAM technique can be used with a focus on improving localization accuracy in GPS degraded environments, by fusing the outputs of a SLAM algorithm with GPS and IMU data [? ]. Or as presented in [? ], the solution focuses on building highly accurate maps online of building interiors or even outdoors, but as a byproduct relative position becomes more accurate as well. The above mentioned papers present a mapping approach that uses industrial-grade hexacopter and octocopter and the same precise LIDAR scanner that can measure up to 100m.

Localization and mapping of the environment for a small quadcopters can be challenging, because adding any extra equipment means extra weight that shortens flight time and changes flight behavior and the onboard computation capacity is usually limited. Skydio2 [? ] uses 6 fisheye cameras and processes the data onboard using a GPGPU to track a moving target while avoiding even the smallest branches in dense forests. Rotating planar scanners are the other common choice, that are also used in the above mentioned papers, undoubtedly provide highly accurate scans and detailed maps. These LIDAR scanners are perfect for professional use and require reliable octo- or hexacopters to carry. The popular Velodyne VLP-16 for example weighs 830g. Adding it to a quadcopter that is under 1kg means significant extra weight and the drone might turn out to be incapable of flying.

By using stationary LIDAR sensors placed around the quadcopter, these unwanted effects can be eliminated and the weight of extra equipment can be kept significantly lower, than using a 3D LIDAR scanner discussed before. This solution might also provide a cheap and computationally light alternative to camera or planar scanner based solutions. The selected single-point sensors need to be placed on a drone in a layout that offers an optimal scan of the environment, therefore the movements of quadcopter during flight needs to be taken into account to be able to scan the environment properly around the vehicle.

In this thesis the building and evaluation of a 3D positioning and mapping system is introduced, using Gazebo simulator. The system uses stationary LIDAR sensors placed on a simulated quadcopter to acquire distance measurements from its environment and uses SLAM technique to map its surrounding, estimate its position and produce an indoor map.

In chapter 2 basic aviation terminologies are introduced that will be used in the thesis, followed by introduction of SLAM techniques and similar products on the market. Visual SLAM techniques describe the use of monocular cameras, stereo cameras or RGB-D cameras to be used for data acquisition, while ORB-SLAM and LSD-SLAM are two popular SLAM techniques for processing data coming from these sensors. Ranging sensors are non-visual sensors that can use sound or light to measure distance, like ultrasonic or LIDAR sensors. Two interesting researches are discussed, that use LIDAR scanners to improve position or map quality online. The algorithm developed by google called Cartographer SLAM is being introduced here. Cartographer is a software package that will be used for SLAM implementation and to evaluate layout design and settings of the stationary sensors. The algorithm will be used with the official ROS wrapper. Lastly all necessary components of the simulation and their task is discussed.

Chapter 3 gives an overview of VL53L1X LIDAR sensor produced by STMicroelectronics and describes the relevant operating modes to be tested and the methodology of the measurement. These results will be used to accurately simulate the sensors. In this chapter initial sensor layouts are proposed for both 2D and 3D SLAM operations.

To be able to compare the performance of different SLAM algorithms, I chose to use the same data for the evaluation of different setups and filter the LIDAR measurements to simulate different layouts. To do so, two LIDAR sensors are placed on the simulated quadcopter to scan the whole sphere of the surroundings of the drone. Data published to ROS during flight is collected into a rosbag, that contains the relevant measurements, that are high resolution LIDAR ranges and IMU data. The LIDAR measurements inside this rosbag is then filtered to simulate a certain number of VL53L1X sensors. Update rate, field of view and maximum ranging distance parameters from the planned sensor measurements will be used to simulate a sensor. The filtered data is fed to Cartographer SLAM that produces the map after proper tuning.

Chapter 4 describes the results of the VL53L1X measurements that was planned in previous chapter. VL53L1X LIDAR has an adjustable region of interest inside the field of view and by moving this region around the field of view, higher resolutions can be achieved from 1x1 to 4x4. It is yet unknown how different parameters affect the SLAM performance, therefore these are introduced incrementally.

For 2D SLAM evaluation the same layout is used in all steps, 13 sensors evenly distributed on the horizontal plane. At first the effects of different resolutions from 4x4 to 1x1 are evaluated, then the effects of lowered sampling rate and lastly the effects of maximum ranging distance. The performance of each setup is determined by calculating RMS error between the estimated trajectory and the ground-truth trajectory, extracted from the unfiltered dataset.

The three introduced parameters are according to results of VL53L1X measurements described in the beginning of this chapter. Using the optimal settings, the number of sensors is decreased until the lowest number if found, that can still provide a reliable SLAM performance. Instead of decreasing the number of sensors by one at each iteration, I chose to do logarithmic search.

The above described evaluation methodology is used for 2D SLAM and the optimal LIDAR settings are used for the evaluation of 3D SLAM. Just like in case of 2D, the sensor parameters are introduced one by one to see their effects independently from each other. Lastly the number of sensors is decreased until the lowest number is found.

# Chapter 2

## Literature review

### 2.1 Basic aviation terminologies

#### 2.1.1 UAV or Drone

UAV is short for Unmanned Aerial Vehicle, in other words an airborne vehicle that is capable of movement without having a pilot on board. Drone is a subset of UAV, a vehicle that is capable of autonomous flight. Usually UAV and the word Drone are used interchangeably in the literature, and it is used so in this thesis.

#### 2.1.2 Body axes used in aviation

In aviation Euler angles are used to describe the orientation of an aircraft. Euler angles are three angles that describe the orientation of a rigid body with respect to a fixed coordinate system. These axes are referred to as Yaw, Pitch and Roll. Rotation around these axes are called Yaw, Pitch and Roll angles respectively. All three angles follow the right-hand rule.

Magnetic north is used as reference for Yaw, so an angle of  $0^\circ$  or  $360^\circ$  means that the vehicle is heading North,  $90^\circ$  means it's heading East. Yaw is sometimes called Heading.

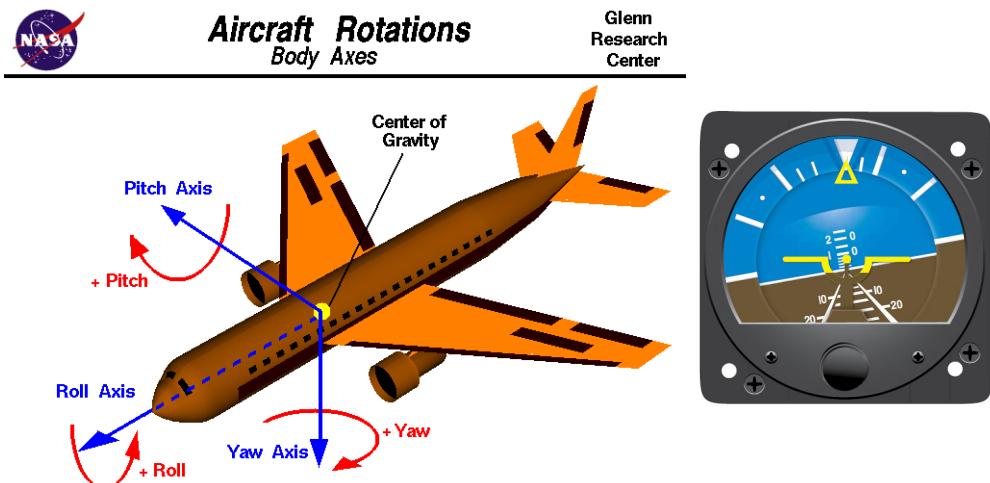


Figure 2.1: Aircraft body axes[?] and attitude instrument

Pitch and Roll angles are calculated in reference to the horizontal plane. The normal vector of the horizontal plane is used for the calculations, and it is the norm of the gravity vector. It is favorable to use the gravity vector, because its direction can be measured using an accelerometer. Both Roll and Pitch angles are measured from -90° to 90°. A Pitch of 90° is straight up and 0° is Horizon. If an aircraft flies with 0° Roll angle, the vehicle is horizontal, on the other hand a 90° Roll angle means the vehicle is turning right and is perpendicular to the horizon. Pitch is sometimes referred to as Tilt.

The limitation of using Euler angles is reached when Pitch or Roll angles approach 90°, because in this case one of these angles become parallel with the gravity vector and the other angle cannot be determined, due to lack of reference. Euler angles and singularities are well described in [? ]. To avoid singularities of Euler angles, Unit quaternions can be used. Quaternions are mainly used for calculations, while Euler angles are used to provide humanly readable values.

### 2.1.3 Ground Control Station

A software running on a ground computer, used for receiving in-flight information via telemetry from a UAV. It displays status and progress of mission, that often includes sensor or video data. It can also be used for sending commands up to the UAV during flight.

Ground Control Station is often referred to as GCS.

### 2.1.4 Multicopters

Multicopter or Multirotor is a generic term to describe a UAV with more than two rotors. The term covers quadcopters, octocopters, hexacopters etc.

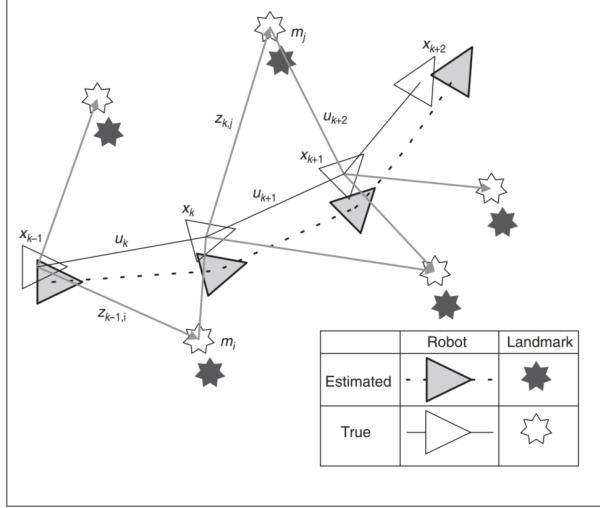
Quadcopters are quickly gaining popularity thanks to their easy usability, and because of commercially available drones that can be used for filming. Octocopters have eight rotors in total and mostly used by professionals. It provides a more stable flight, can lift more weight and it is safer, because it can fly with any of the rotors damaged.

## 2.2 Simultaneous Localization And Mapping - SLAM

Simultaneous Localization And Mapping is the computational problem that asks if it is possible to place a robot in an unknown environment and for the robot to incrementally build a consistent map of this environment while simultaneously determining its position. A good overview of SLAM is presented in[? ] and [? ] focusing on finding a solution to the above proposed problem in general, with no dependency on the type of sensor to be used.

During SLAM a robot is placed in an unknown location and it is capable of estimating its trajectory and location of all landmarks online, without the need of priori knowledge of the location. The robot makes relative observations of its surroundings, a number of unknown landmarks using sensors as seen on figure 2.2. The following quantities are defined at a time instant  $k$ :

- $\mathbf{x}_k$ : State vector describing the location and orientation of the robot
- $\mathbf{u}_k$ : Control vector, applied at  $k-1$  and drove the robot to state  $\mathbf{x}_k$  at time  $k$



**Figure 2.2:** The essential SLAM problem presented in [? ]

- $\mathbf{m}_i$ : Landmark location vector. A time independent vector that describes the location of a static landmark
- $\mathbf{z}_{ik}$ : Observation taken from the robot to the  $i$ th landmark at time  $k$
- $\mathbf{X}_{0:k} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$  A set of all robot location vectors until time  $k$
- $\mathbf{U}_{0:k} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$  A set of history of all control vectors until time  $k$
- $\mathbf{m} = \{\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_n\}$  A set of all landmarks
- $\mathbf{Z}_{0:k} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k\}$ : A set of all landmark observations until time  $k$

In probabilistic form of SLAM, the probability distribution function 2.1 needs to be computed for all times  $k$ . In other words the current state vector and the landmark location probabilities need to be calculated based on all  $k$  observations, control vectors and the initial state of the robot.

$$P(\mathbf{x}_k, \mathbf{m} | \mathbf{Z}_{0:k}, \mathbf{U}_{0:k}, \mathbf{x}_0) \quad (2.1)$$

A recursive solution is desirable for the SLAM problem. Starting with an estimate for the posteriori distribution (equation 2.2) at time  $k-1$ , then a control vector  $\mathbf{u}_k$  is used to estimate the next state and observations using Bayes theorem. This calculation requires an observation model and a state transition model.

$$P(\mathbf{x}_{k-1}, \mathbf{m} | \mathbf{Z}_{0:k-1}, \mathbf{U}_{0:k-1}, \mathbf{x}_0) \quad (2.2)$$

The observation model 2.3 describes the probability of making an observation  $\mathbf{z}_k$  given the current robot and landmark locations.

$$P(\mathbf{z}_k | \mathbf{x}_k, \mathbf{m}) \quad (2.3)$$

The state transition model 2.4 describes the probability of the next position of the robot based on the previous state and the control vector at time  $k$ . The state transition is assumed to be a Markov process, so the next state  $\mathbf{x}_k$  depends only on the current state

$\mathbf{x}_{k-1}$  applied control  $\mathbf{u}_k$ .  $\mathbf{x}_k$  is also independent from both landmark locations  $\mathbf{m}$  and observations  $\mathbf{z}_k$ .

$$P(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k) \quad (2.4)$$

Using equations 2.2, 2.3 and 2.4, the SLAM algorithm can now be described using a two-step recursive form: prediction and correction. The prediction step or time-update is used for estimating the next state and map, based on all posteriori observations, control vectors and initial state.

$$P(\mathbf{x}_k, \mathbf{m} \mid \mathbf{Z}_{0:k-1}, \mathbf{U}_{0:k}, \mathbf{x}_0) = \int P(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k) \times P(\mathbf{x}_{k-1}, \mathbf{m} \mid \mathbf{Z}_{0:k-1}, \mathbf{U}_{0:k-1}, \mathbf{x}_0) d\mathbf{x}_{k-1} \quad (2.5)$$

The measurement update provides a correction to the prediction state, using observations made at time  $k$ .

$$P(\mathbf{x}_k, \mathbf{m} \mid \mathbf{Z}_{0:k}, \mathbf{U}_{0:k}, \mathbf{x}_0) = \frac{P(\mathbf{z}_k \mid \mathbf{x}_k, \mathbf{m}) P(\mathbf{x}_k, \mathbf{m} \mid \mathbf{Z}_{0:k-1}, \mathbf{U}_{0:k}, \mathbf{x}_0)}{P(\mathbf{z}_k \mid \mathbf{Z}_{0:k-1}, \mathbf{U}_{0:k})} \quad (2.6)$$

The localization problem can be solved with the assumption the map is known:

$$P(\mathbf{x}_k \mid \mathbf{Z}_{0:k}, \mathbf{U}_{0:k}, \mathbf{m}) \quad (2.7)$$

And the mapping problem can be solved with the assumption the location is known:

$$P(\mathbf{m} \mid \mathbf{X}_{0:k}, \mathbf{Z}_{0:k}, \mathbf{U}_{0:k}, ) \quad (2.8)$$

These problems introduced in 2.7 and 2.8 are dependent on each other and need to be solved simultaneously. In the early stages of SLAM mapping and localization were tried to be solved independently and work often focused on either of them. Once the realization came that combined mapping and localization can be formulated as a single estimation problem, the problem became convergent.

Many solutions can be found for probabilistic SLAM problem, the most common is the use of the extended Kalman filter (EKF) that provides an optimal estimation to non-linear models with additive Gaussian noise. The motion model can also be represented with a non-Gaussian probability distribution, that leads to the use of the Rao-Blackwellized particle filter, or FastSLAM algorithms.

The above introduced SLAM problem was a general introduction, it is not dependent on the sensor used for observations. Many kind of observers can be used for this purpose, but the two main categories are camera based and ranging sensor based. Visual observers mostly use simple or special camera configurations for depth sensing, while ranging sensor based observers use a physical phenomena to measure distances like sonars or LIDARS.

### 2.2.1 Visual SLAM

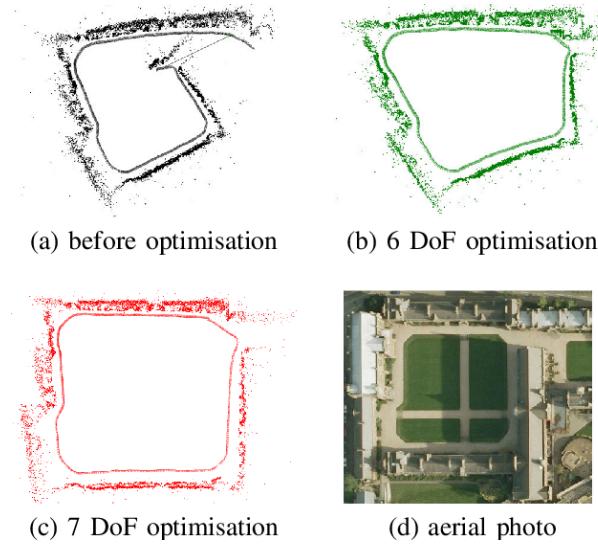
Visual SLAM is a subset of SLAM algorithms that processes camera images to determine the camera's position and orientation relative to its surroundings, while building a map of its environment. In the simplest case a monocular camera is used to collect images,

but more advanced active or passive stereo cameras and RGB-D cameras are also used. A pro of using images for SLAM is that images contain high amounts of data, therefore very detailed maps can be made. On the other hand, processing big amounts of data means higher complexity and demands higher processing capabilities.

Two main categories of visual techniques are feature-based and direct SLAM algorithms. Feature-based techniques detect certain key-points called features in images, like corners and edges, and only use these features to extract depth information. Direct SLAM algorithms however don't search for features, but use the regions of high intensities on the image to estimate the location and surroundings. Two popular visual SLAM methods are feature-based ORB-SLAM and direct LSD-SLAM.

### 2.2.1.1 Monocular camera

The pro of using monocular camera for localization and mapping is that each image contains high density of information and cameras will always be cheaper easier to use than other depth sensing camera rigs. A big pro, but in the same time a big con of using a monocular camera for SLAM is that a single camera cannot detect the scale. It is an advantage, because the same camera can be used indoor on a small scale even down to centimeters, and used outdoors on a scale of kilometers. Nonetheless during transmission between these scales, the scale can drift and therefore size of landmarks will not be presented correctly. This is called scale-drift, that algorithms try to compensate.

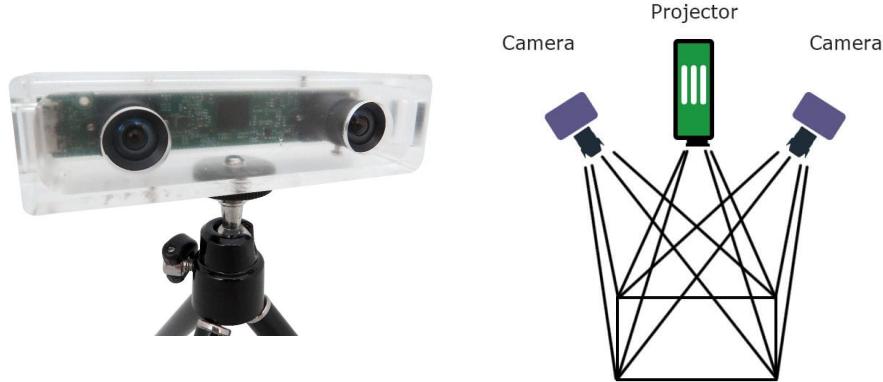


**Figure 2.3:** Scale drift and compensation illustration from [? ]

### 2.2.1.2 Stereo camera

A stereo camera is a type of camera with typically two image sensors. Two image sensors mimic the human binocular vision and makes depth perception possible. The two cameras are placed relatively close together so there is a high correlation between the images taken by these cameras. To find correlations between the two images sufficient details and structure needs to be present on the images, therefore light conditions affect the performance of depth-sensing. This technique is also called passive stereo vision.

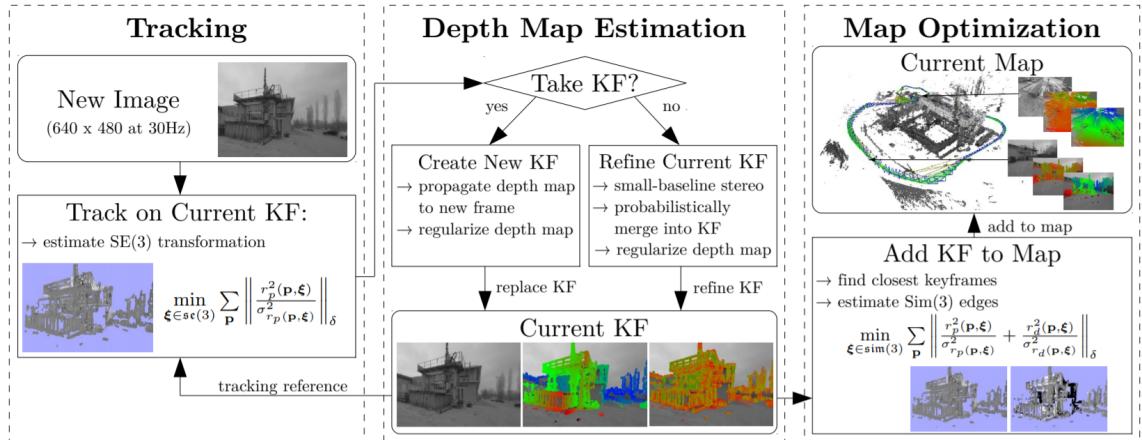
Active stereo cameras solve this problem employing a projector onboard, that projects a structured light to simplify the stereo matching problem. Using this technique the depth sensing performance is less affected by lighting conditions and reoccurring patterns will not confuse the matching of the images.



**Figure 2.4:** Passive stereo camera on the left, concept of active stereo camera on the right

### 2.2.1.3 LSD-SLAM

LSD-SLAM for monocular cameras were proposed in [? ] and for stereo or RGB-D cameras in [? ]. LSD stands for Large-Scale Direct SLAM and it can be seen by the name that it is a direct SLAM algorithm. The map of the surroundings is created based on keyframes that consists of three fields, a camera image, an inverse depth map and the variance of the inverse depth map. The solution of the problem proposed by SLAM is divided into three parts: tracking, depth map estimation and map optimization as seen on figure 2.5.



**Figure 2.5:** LSD-SLAM overview of the complete algorithm [? ]

New images are input during the tracking phase and the position is estimated with respect to the current keyframe from previous iteration. The depth map estimation component uses tracked frames to either refine or replace the current keyframe. If the camera has moved more than a set threshold, a new keyframe is initialized with the combination of current and new keyframe. If the tracked keyframe is not taken, then it is used to refine the

current keyframe by performing small-baseline stereo comparisons and regularization of the depth map. Once the current keyframe is updated, it is built into the global map, by the map optimization. Keyframes that are close to each other can be detected and closed, this process is called loop-closure. For loop closures and scale-drift compensation, similarity transform is used. The map is further optimized using pose graph optimization from g2o package.

#### 2.2.1.4 ORB-SLAM

ORB-SLAM is a feature-based SLAM algorithm introduced in [? ]. The features extracted are represented by ORBs, therefore the name of the algorithm. ORB-SLAM has three threads that run in parallel: tracking, local mapping and loop closing.

The tracking phase is responsible for localizing the camera and to decide when to insert a new keyframe. The features extracted from the new image are FAST corners, where FAST is a computationally efficient process to find corners on images. These corners are described using ORB, that is a method to describe FAST corners by a center of mass and an orientation parameter. This way the place and orientation is known of each feature. The initial pose is estimated using a constant velocity model and if the tracking is lost, relocation is done instead of pose estimation. Track local map is a map of keyframes that share a common map point. Finally it is decided if new keyframe is need to be created.

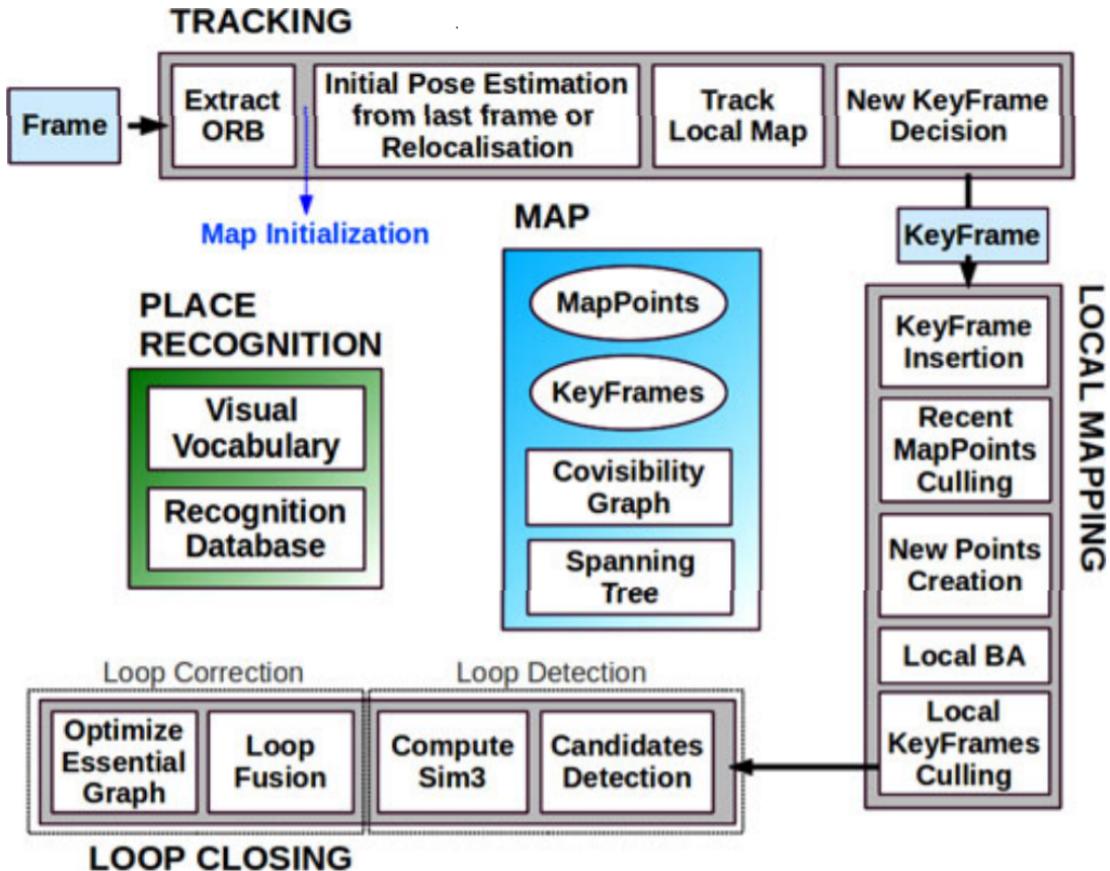


Figure 2.6: ORB-SLAM overview [? ]

In the first step of local mapping the new keyframe is inserted into the co-visibility graph and it is linked keyframe to keyframe with the most points in common. ORB features from

the connected keyframes are used to calculate new map points using triangulation. Unmatched ORBs are compared with other unmatched ORBs of other keyframes. Projection error and consistency is checked, then the map is projected to the matched keyframes. The new map points need to be validated, by going through a test to increase the likelihood that these map points are valid part of the map. In local bundle adjustment step, the current keyframe and connected keyframes through the co-visibility graph are optimized. Finally abundant keyframes get discarded to keep simplicity.

Loop closing is quite effective in ORB-SLAM, because matching ORBs can be done with relatively high confidence. Similarity transformation is calculated.

### 2.2.2 Ranging sensor based SLAM

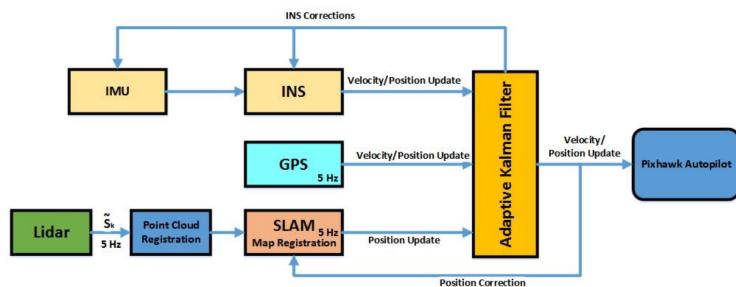
Cameras are cheap, popular and can be used to create highly detailed maps, but they produce high amounts of data that demands high processing capabilities. Ranging sensor based approaches produce lower amount of data that is dependent on the scale, these sensors have a minimum and maximum range they can measure in between.

Ultrasonic sensors have been used for distance measurements for many years, but they suffer from significant measurement noise, that comes from background noise and the variability of speed of sound. Almost no development has been seen on ultrasonic sensors in the previous years LIDAR sensors on the other hand are actively developed, because this technique uses light to measure distance, offer more accurate measurements, typically lower form factor, higher resolution and update rates. LIDAR sensors are also significantly more expensive than most of ultrasonic sensors and camera based SLAM systems.

A planar laser scanner is a LIDAR sensor that is extended to make distance measurements in 2D. Typically it is used for scanning the horizontal plane. Usually the scanning is solved by rotating a LIDAR sensor in  $360^\circ$  while it is making measurements evenly distributed on the plane. Planar scanners are good for building floor maps and navigating in indoor environments. Some scanners are further extended for 3D scanning.

#### 2.2.2.1 3D LIDAR SLAM to improve localization accuracy

The study of University of California[?] presents an algorithm that fuses LIDAR, IMU and GPS signals to have a more accurate estimation of the absolute position and velocity of a UAV. The LIDAR scanner provides local position updates using SLAM technique,



**Figure 2.7:** 3D LiDAR SLAM Integration with GPS/INS for UAVs in Urban GPS-Degraded Environments [?] - algorithm overview

GPS provides corrections when available and an Internal Navigation System is used as an

additional input to the Adaptive Extended Kalman filter. The outline of the filter can be seen on figure 2.7.

The 3D LIDAR scanner used is Velodyne VLP-16, that has a vertical field of view of  $30^\circ$  with a vertical resolution of  $2^\circ$ . On the horizontal axis the resolution is  $0.1\text{-}0.4^\circ$  and rotation rate can be adjusted between 5-20Hz. The sensor weighs 830g according to the company website.

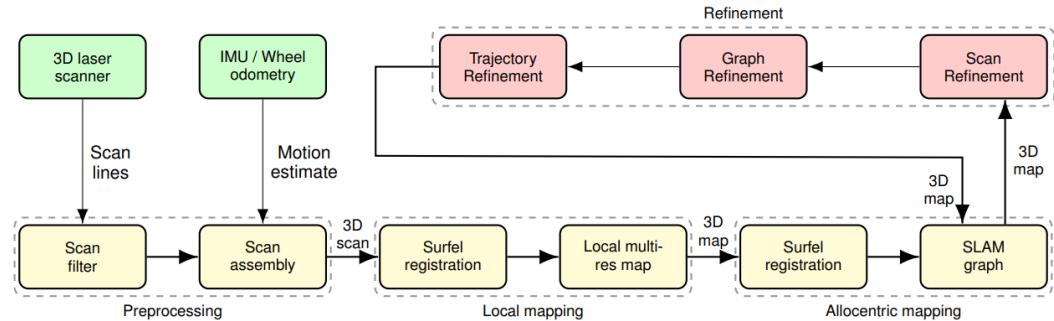
During the experiment an octocopter is used to carry the sensor. Octocopters are stable and can carry high amounts of weight. The proposed filter is capable of reducing the GPS drift of 24.3m and LIDAR drift of 7.5m to 3.42m. This shows that fusing GPS, LIDAR and IMU measurements result in a more accurate position estimate in GPS-degraded environments. The capabilities of the LIDAR was also tested on an even moon-like surface, with very few landmarks that can be used by the SLAM algorithm. In this case GPS data is much more accurate than LIDAR SLAM, therefore LIDAR data is not used for calculation of position updates in these cases.



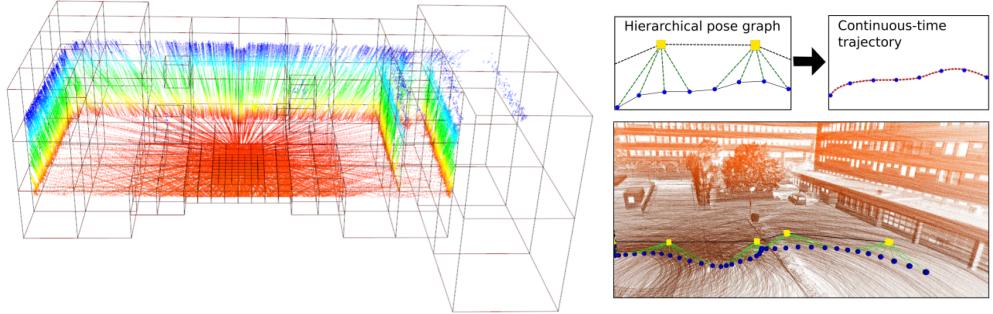
**Figure 2.8:** GPS, LIDAR and EKF position estimates and position drift after 405 meters[? ]

### 2.2.2.2 3D LIDAR SLAM for mapping

The solution seen in 2.2.2.1 uses a LIDAR scanner to improve position and velocity data, by fusing the output of SLAM algorithm with sensors. The map created simultaneously is not used in that solution. 3D LIDAR data and SLAM can be also used to create highly detailed maps and this paper presented by David Droeßel and Sven Behnke [? ], discusses an efficient implementation of SLAM with focus on improving online mapping quality.

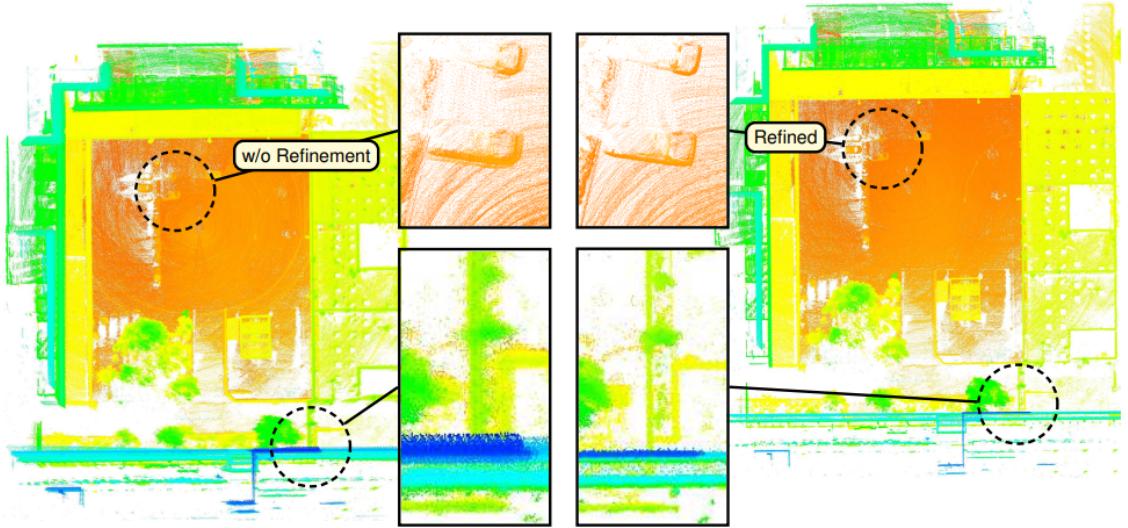


**Figure 2.9:** Efficient Continuous-time SLAM for 3D LIDAR-based online mapping[? ] - algorithm overview



**Figure 2.10:** Grid-map presented in [? ] on the left, estimated trajectory and built map on the right

The system first aggregates range measurements into a robot-centric grid-map [? ], that has high resolution in the close proximity and lower resolution with increasing distance. A grid cell stores the cell's occupancy probability and a surface element(surfel), that describes the mean of samples and covariance. IMU or odometry measurements are used to account for motion of the sensor during acquisition, to compensate rolling shutter artifacts. After the preprocessing step, surfels become registered and the local map is updated. The older local map measurements are overwritten with the current 3D scan. During allocentric mapping phase, the 3D local multiresolution maps from different view poses are aligned with each other using surfel based registration. Local maps that are registered to the SLAM graph are subject to refinement. During these steps, the scans are refined based on available local maps, then graph and sensor trajectories are also refined. Local mapping, allocentric mapping and refinement phases are independent from each other, so these can run parallel.



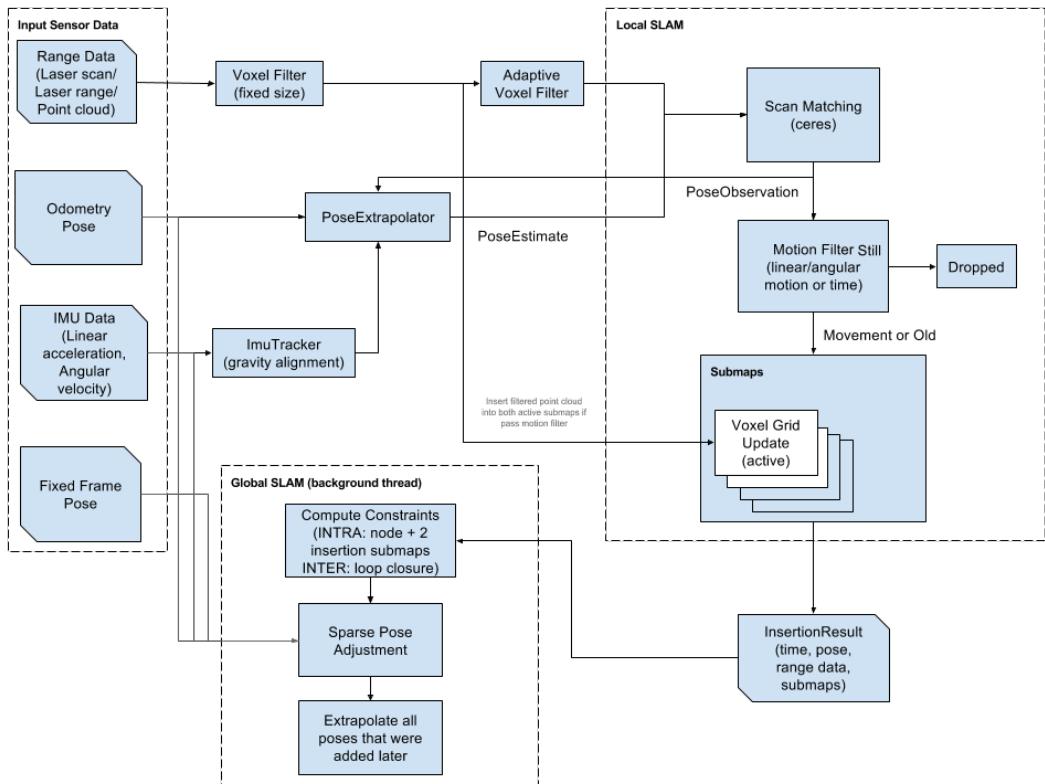
**Figure 2.11:** Comparison of maps with and without refinement[? ]

### 2.2.3 Cartographer SLAM

Cartographer is a system developed by Google that provides real-time simultaneous localization and mapping in 2D and 3D across multiple platforms and sensor configurations.

Cartographer is used by Google street view for mapping of building interiors, using a backpack mounted LIDAR scanner.

There are officially available SLAM packages in ROS, the most popular packages are Gmapping, Cartographer and Hector SLAM. For the scope of this thesis, only ROS compatible packages are being concerned. Gmapping and Hector SLAM are popular choices for SLAM among developers and even unofficial 3D extensions are available, but officially they only support 2D mapping. Cartographer supports 3D SLAM without any modifications and for this reason Cartographer has been selected. For some yet unforeseen if Cartographer is proved to be incapable of handling the selected sensors, other SLAM algorithms can be considered.



**Figure 2.12:** Cartographer SLAM overview[? ]

On the other hand Cartographer ROS package supports both 2D and 3D SLAM and has a great and active community. The overview of Cartographer is seen on figure 2.12. The core components are sensor inputs, Local SLAM and Global SLAM. On a higher abstraction the job of local SLAM is to build submaps and the job of global SLAM is to tie submaps the most consistently together. It is common to build submaps or local maps in SLAM implementations, because by updating the whole map increases complexity quadratically, but by using local maps or submaps the complexity can be limited to the contents of the submap that is preferable during online processing. In this case alignment of submaps needs to be solved, but it requires less processing and it can be run in parallel.

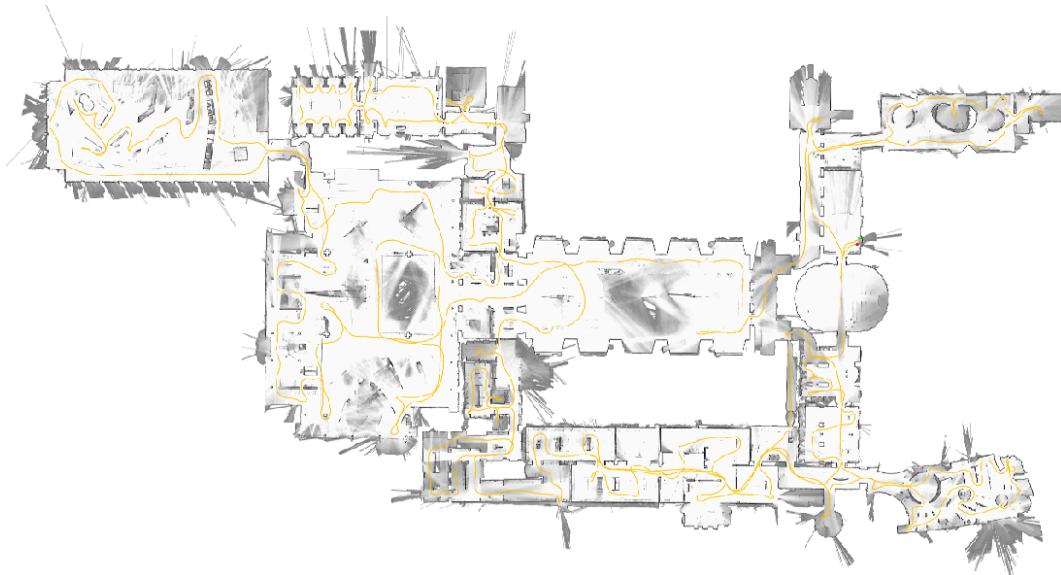
The most important input of Cartographer is the Range Data, range measurements coming from LIDAR sensors. The data is first pre-filtered, because some measurements are irrelevant, the sensor might be directed to a part of the robot or it can be covered by dust, and some sensors set unsuccessful range measurements to a value that is significantly higher than the maximum range of the sensor. These outliers need to be filtered

and Cartographer uses a bandpass filter for pre-filtering, that keeps the values in a pre-defined range. The minimum and maximum values need to be set according to the sensor specifications.

Close objects are very often hit by the LIDAR measurements and offer more points, however distant objects offer much less points. The Voxel filter down-samples raw points where density is higher, but leaves scarce measurements to reduce computation needed, while keeping data integrity.

Inertial Measurement Units (IMU) provide a direction of gravity vector and a noisy estimation of the robot's movement. IMU data has to be provided for 3D SLAM because it greatly reduces complexity of scan matching, while IMU data is optional for 2D SLAM. Odometry Pose and Fixed Frame Pose are optional inputs of Cartographer, that can further reduce complexity.

In Local SLAM phase, first scans are inserted into a submap at the best estimated position. Scan matching is done only against current submap and not against the global map. During this process the error of pose estimates accumulate in the world frame. To reduce the accumulated error over time, pose optimization is run regularly. All finished submaps are considered for loop closure, that happens on a background thread Global SLAM.



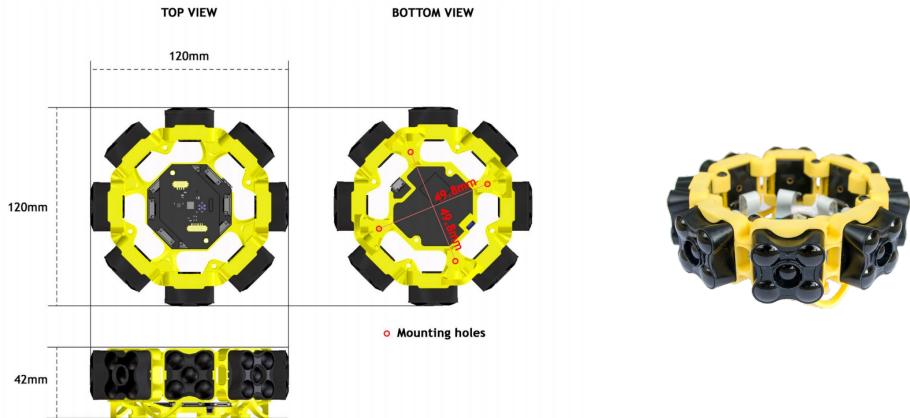
**Figure 2.13:** Cartographer SLAM example[? ]

## 2.3 Similar products available on the market

### 2.3.1 Terabee TeraRanger Tower

TeraBee offers an off-the-shelf solid-state LIDAR system with the purpose of collision avoidance for drones. In their solution 8 sensors are evenly distributed around the vertical axis with a controller board in the middle. TeraRanger's interface is compatible with Pixhawk 4 flight controller board and PX4 flight controller software, that makes integration easy into systems based on these components.

Each block of the array is a standalone LIDAR sensor, that can be used separately and supports different mount configurations. The company offers a long-range and fast-ranging



**Figure 2.14:** Terabee TeraRanger Tower Evo dimensions

version of these sensors and depending on the type an update rate of 320 Hz can be achieved. The fact that system comes with a ROS package and a  $2^\circ$  field of view makes it a potential candidate for indoor mapping and positioning in two dimensions. The price of this setup starts from 599 euro[? ].

	Long-range	Fast-ranging
Range	0.5m up to 60m	0.75m up to 8m
Update rate	120Hz/sensor	320Hz/sensor
Field of View	$2^\circ$	$2^\circ$
Accuracy	$\pm 4cm$ in the first 14m, 1.5% above	$\pm 12cm$

**Table 2.1:** TeraRanger Tower Evo specifications

Terabee provides a tutorial video on their website[?] how to connect the sensor array to a UAV that uses Pixhawk 4 flight controller and the process of configuration in two different GCS programs. I have learned that PX4 flight stack supports lidar measurements for obstacle avoidance and it can be configured from a GCS program. It seems a reasonable choice to integrate this feature into such product.

### 2.3.2 Crazyflie Multi-ranger deck

The company Bitcraze has developed a mini quadcopter mainly for educational purposes. The current version is called Crazyflie 2.1 and measures only 92x92mm with a height of 29mm and a weighs 27g. Extra sensors and peripherals can be attached to the top of the quadcopter using extension boards.

The extension board called Multi-ranger deck has 5 VL53L1X sensors by STMicroelectronics facing forwards, backwards, left, right and up. This project is similar to the product of Terabee described in 2.3.1, but in a smaller size factor and with significantly lower weight.

An introduction video can be found on the product website [? ], where a SLAM algorithm is used to create a map and localize the drone. This serves as a proof of concept, that stationary VL53L1X sensors can be used for mapping and positioning in two dimensions. The company provided no information of the SLAM algorithm used or from the quality of the map.

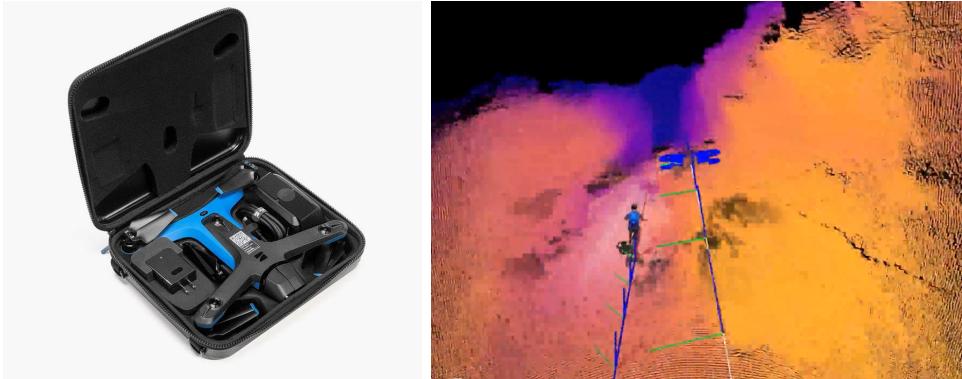


**Figure 2.15:** Crazyflie Multi-ranger deck, SLAM example project

### 2.3.3 Skydio 2

Skydio 2 is a quadcopter built for autonomous tracking for filming purposes. This drone was released in November 2019 soon after I have started working on this thesis. It is equipped with 6 wide angle cameras besides the main camera that capture the vehicle's environment with 500Hz update rate and process it in real time onboard. [? ]

The drone uses the data coming from the cameras to avoid obstacles and to follow the target that is being tracked. Based on reviews, the drone's obstacle avoidance is highly advanced and it is even able to follow a downhill biker even in dense forests.



**Figure 2.16:** Skydio2 autonomous drone and its surroundings using 6 wide angle cameras

The drone is controlled by the user's phone, a Beacon or a traditional radio controller. When using a phone the drone purely relies on the camera feed for tracking. This only works with constant line of sight, otherwise the drone might get stuck at some point and requires user intervention. Besides line of sight a Beacon can be used that improves the tracking performance.

### 2.3.4 Velodyne Velabit

Velabit is an affordable directional LIDAR sensor developed by Velodyne. The sensor will be available for customers in mid-2020 and only official specifications are available at the time [? ]. The sensor has a  $60^\circ$  horizontal and  $10^\circ$  vertical field of view and it can range up to 100m. It is placed inside a housing with size of 6x6x3.5cm. The weight of this sensor

is currently unknown, but based on it's size, it is expected to be low enough so that it can be equipped even on small sized drones as a precise mapping solution.



**Figure 2.17:** Velodyne Velabit LIDAR sensor[? ]

## 2.4 Components review

Before actually building the system and experimenting on an expensive quadcopter in real world environment, it is cheaper, safer and easier to do tests in a simulated environment.

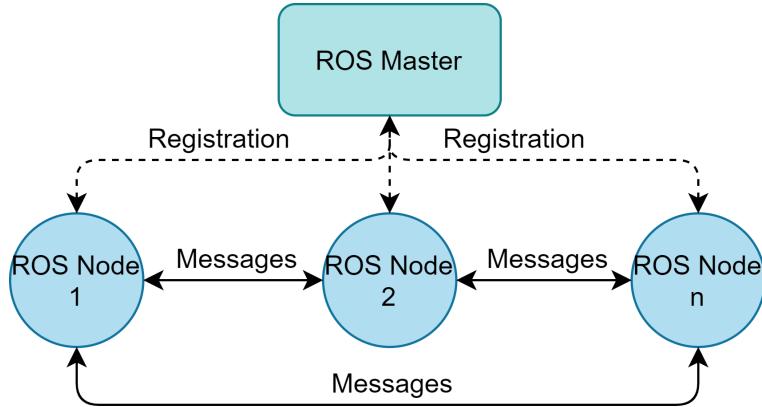
The PX4 Firmware repository[? ] contains Gazebo simulation files to give developers a head start in simulation of their product. Upon these simulation quadcopter models LIDAR sensors can be placed and moved around with ease, without the need of wiring or external infrastructure. This simulation is suitable for Software in the loop (SITL) testing, meaning that the same software developed for the simulation can be used on a real quadcopter with little or no modifications. As for communication of the software components Robot Operating System will be used.

### 2.4.1 Robot Operating System

Robot Operating system (ROS) is a set of software libraries and tools that help developers to build robust general-purpose robot applications[? ]. ROS provides hardware abstraction of the underlying robot, generalizes the interfacing of these robots and allows simple high-level usage.

The core component of ROS is a public-subscribe communication protocol similar to MQTT protocol used on the field of IoT. The ROS system is comprised of a number of independent communicating parties called nodes. Each node can subscribe to multiple topics and receive a stream of messages published to these topics by other nodes. For example a temperature filter node might subscribe to a topic of "/temperature\_sensor" and publish the filtered temperature values to a topic called "/filtered\_temperature\_sensor". The benefit using this solution is that the nodes achieve communication exclusively via this protocol so each node lives independently from another. Nodes in ROS do not need to be on the same system or even on the same architecture. Nodes can be run on different computers or even on microcontrollers or smartphones making the system flexible.

To start a ROS session, a ROS Master needs to be started first. After starting the master, nodes can be started and each one registers the topics it wants to subscribe/publish to. The master handles these requests and helps the nodes to establish connection between each other so communication can be started.



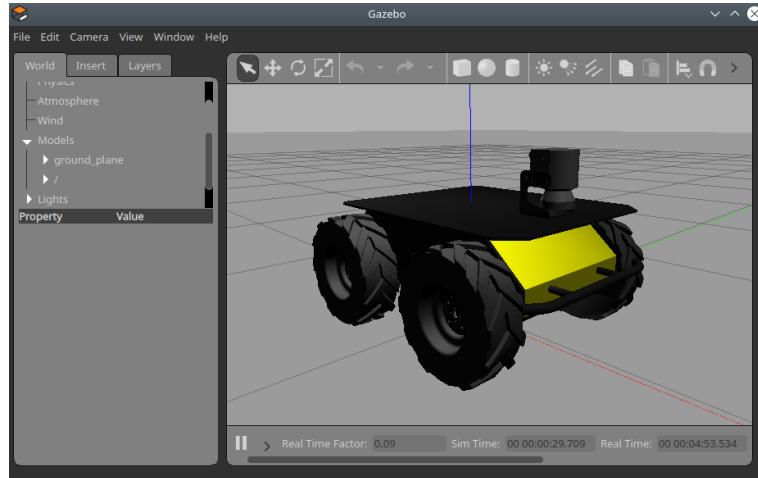
**Figure 2.18:** ROS publish subscribe communication

ROS is language independent so nodes can be implemented in different programming languages. While C++ is the most common language for product development, official Python wrappers can be used for fast prototyping.

#### 2.4.2 Gazebo Simulator

Gazebo[?] is an open-source 3D simulator for robotics, with the ability to accurately simulate robots in complex indoor or outdoor environments. It is similar to game engines, but produces more realistic, physically correct behavior. Gazebo is free, open-source, actively developed and has gained high popularity during the last years.

Gazebo comes with a growing number of robots and sensors. For simulations one can choose from the officially available robots or create a custom robot using SDF files. Any of these robots can be customized and any number of sensors can be added. Using simulations sensor data can be generated and development can be started without having the actual hardware.



**Figure 2.19:** Gazebo ROS Husky robot demo

Why Gazebo is particularly interesting, is because it can be started in a ROS compatible mode. Gazebo started with ROS wrapper will actively channel all sensor readings, states and many more to ROS topics, allowing nodes outside of the simulation to subscribe and publish to these topics. Using this technique nodes can be developed in a truly

hardware independent manner. The simulation can be replaced by a real robot and the nodes developed during this phase can be used without changes on the real hardware. For example a Husky robot can be simulated in Gazebo and controlled through ROS. The Gazebo simulation of a Husky can be seen on figure 2.19.

### 2.4.3 PX4 Autopilot Software

PX4 is a powerful open-source autopilot flight stack. PX4 can control many different vehicle frames and vehicle types, supports wide variety of sensors and peripherals. The project is a part of Dronecode, a nonprofit foundation working on open-source projects to achieve standardization of drones.[? ]

The core component of PX4 is the flight stack or flight controller that controls the motors based on sensor measurements and executes commands received from ground control. Besides the flight stack the repository also contains Gazebo simulation files of multiple vehicles and an extensive `Makefile` that makes starting all components of the simulations relatively easy.

#### 2.4.3.1 PX4 simulation using MAVROS

The logical units of the simulation can be seen on figure 2.20. By executing a `make` command the Gazebo Simulator, the PX4 SITL flight stack and the API/Offboard units are started at once. The PX4 SITL flight stack connects on a TCP port to the simulator, receives sensor data, calculates and sends motor and actuator values to the simulated vehicle. There are two options for the navigation of the simulated drone: using an Offboard API or a Ground Control Station. For autonomous flight control an Offboard API needs to be used and MAVROS is a suitable choice, it basically offers a ROS interface for the drone. It forwards all messages from the flight stack to ROS topics and vice versa.

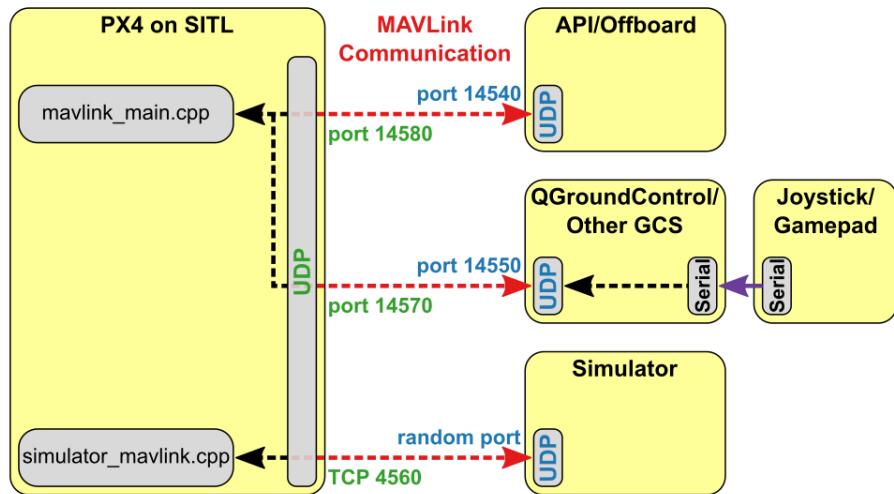
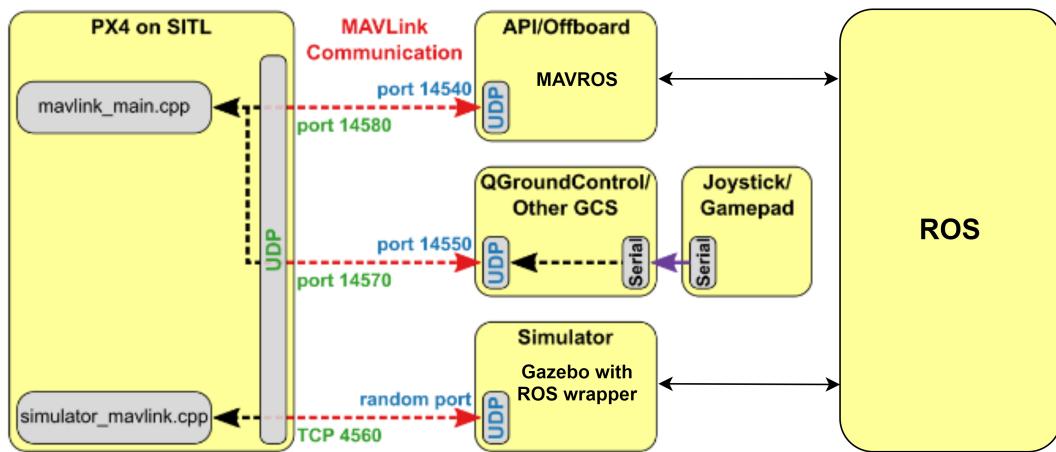


Figure 2.20: PX4 SITL Simulation Outline [? ]

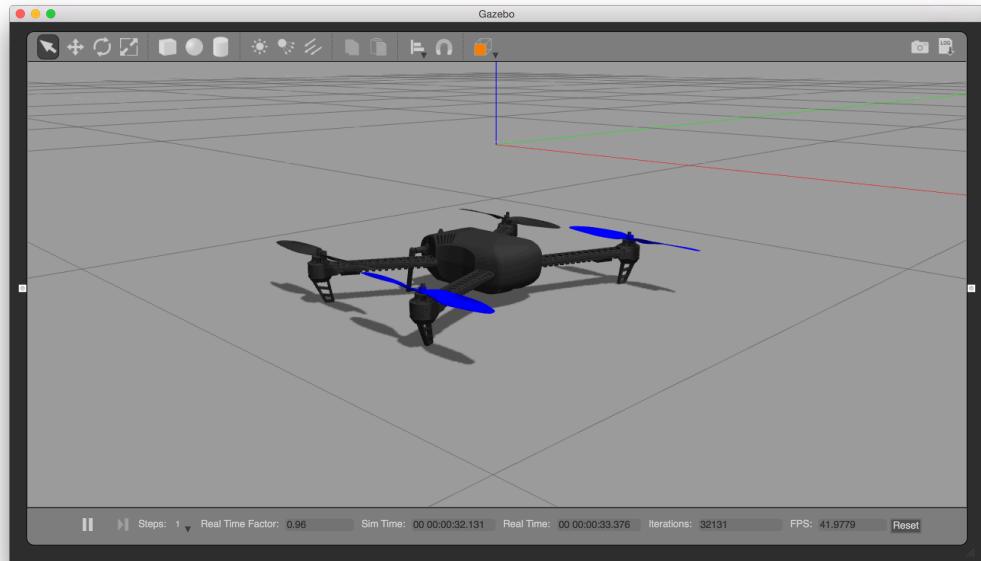
#### 2.4.3.2 PX4 simulation using Gazebo with ROS wrapper

Drone models are described using SDF files that are XML based and by modifying these files new sensors can be added. Gazebo loads the selected vehicle and the sensors, but

if a selected sensor is not supported by PX4, its measurements will not be forwarded to MAVROS and therefore to ROS topics. It can be troublesome to always use sensors that are supported by the flight controller, not to mention that the MAVLink protocol is designed for small messages so the bandwidth can be low in some cases. Luckily Gazebo can be started with a ROS wrapper and therefore all sensors can be accessed on ROS topics. However this technique cannot be used on a real drone with the original PX4 firmware, because it purely relies on MAVLink messages, but for simulation purposes it's an elegant solution. This modified overview of the units used for simulation and a screenshot of an Iris drone simulated in Gazebo can be seen on figure 2.21 and 2.22. MAVROS is still needed to be run, because there are some messages that would not be published to ROS otherwise.



**Figure 2.21:** PX4 Simulation logical unit overview using Gazebo with ROS wrapper



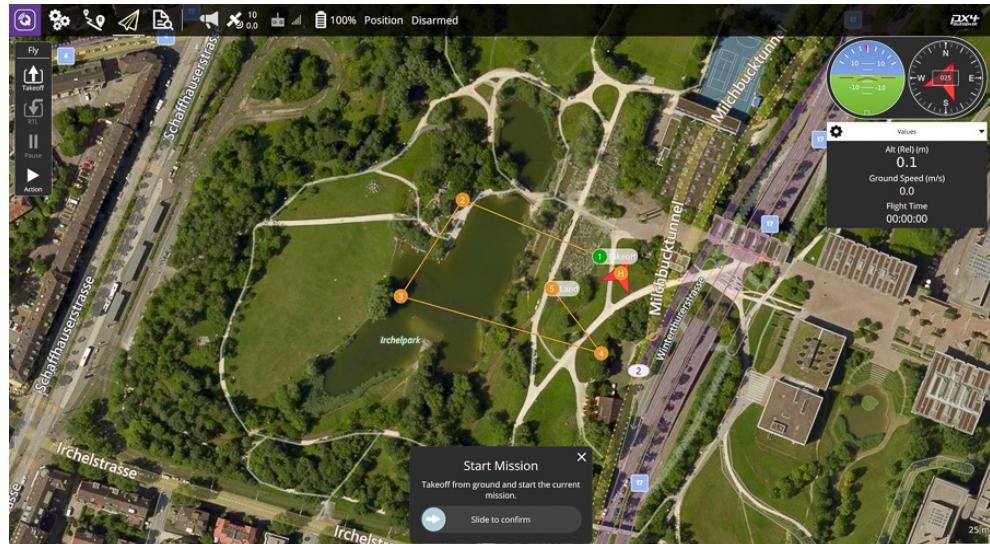
**Figure 2.22:** Iris drone in gazebo

#### 2.4.4 QGroundControl

QGroundControl(QGC) is an open-source ground control station software, developed as a part of Dronecode. It provides full flight control and vehicle setup for PX4 and ArduPilot powered vehicles and for any flight controller that uses MAVLink communication protocol. It provides easy and straightforward usage for beginners, while delivering high end features for experienced users too.

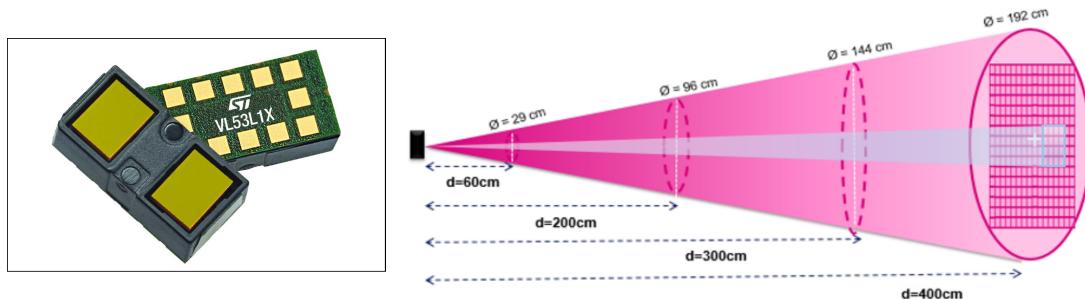
QGC is mainly used for flight controller setup and configuration. When setting up a PX4 autopilot based flight controller, the controller needs to know the type of the vehicle and the frame that is used. Besides selecting vehicle type, internal sensors need to be calibrated and telemetry messages need to be configured. Low level parameters can also be adjusted and obstacle avoidance can also be setup here, as seen in the product of Terabee in 2.3.1, but even low level flight controller parameters can be changed.

QGC can also be used for mission planning for autonomous flight and gives instant feedback of the progress of a mission by visualizing the drone on a map and the most important sensor data.



**Figure 2.23:** QGroundControl main screen[? ]

#### 2.5 VL53L1X LIDAR sensor



**Figure 2.24:** VL53L1X sensor [? ] and Region of Interest [? ]

VL53L1X is described as a long distance ranging Time-of-Flight sensor by STMicroelectronics [? ]. The size of the module is 4.9x2.5x1.56mm, it emits 940nm invisible laser and uses a SPAD(single photon avalanche diode) receiving array with a field of view of 27°. The module runs all digital signal processing on a built in low power microcontroller that frees up the host processor from these tasks. The sensor according to the datasheet is capable of 50Hz ranging frequency and up to 400cm distance measurements. The size of the laser receiving array can be programmatically changed by setting region-of-interest(ROI) size. This way the sensor provides multizone operation and a higher resolution, than by using the whole SPAD array.

I chose to use this sensor, because of its small size, availability in the university laboratory, low hardware infrastructure needs and previous experience.

# Chapter 3

# System design

In this chapter I'm introducing the planned steps to build the previously described system. Due to the recent COVID-19 pandemic I have no access to the university laboratory at the time I'm writing this thesis and therefore the planned system is not built physically.

In order to accurately simulate the selected VL53L1X LIDAR sensor, measurements need to be conducted in different operating modes. Taking into consideration the parameters of the sensor, the number of sensors and the layout can be estimated for both 2D and 3D usage.

The only way to compare the performance of the LIDAR settings, layout and SLAM algorithm is to use the same dataset in every setup. For this reason I have placed a simulated LIDAR scanner on the top of the drone that evenly scans the top hemisphere, and one sensor on the bottom that scans the bottom hemisphere. The complete scan coming from the two sensors cover the whole sphere around the drone with ranges evenly distributed. The drone is flown around inside a building while LIDAR data is recorded into a rosbag file.

The recorded rosbag from the previous step is then filtered to simulate the selected LIDAR layout and settings like maximum distance, sampling rate and field of view. The filtered data is played back to Cartographer SLAM that creates the map and pose estimates for the simulated setup and can be used for evaluation.

To evaluate the performance the RMS error is calculated between the SLAM trajectory and the ground-truth trajectory. Ground-truth trajectory is provided by Gazebo.

## 3.1 VL53L1X measurements

In order to simulate the chosen VL53L1X sensor accurately, experience needs to be gained using the device. There are several factors that affect the quality of the map produced by the SLAM algorithm. Choosing the right settings for the LIDAR sensor therefore is a crucial part that cannot be determined independently from the SLAM algorithm. At this point it is unsure if what resolution, low maximum distance but higher sampling rate is beneficial for SLAM algorithm or the other way around. My approach is to first explore the capabilities of the sensor in all relevant operating modes, measure the ranging performance like update rate, maximum ranging distance and noise level of the measurements. Later on the optimal operating mode can be chosen based on the SLAM performance.

The sensor has three parameters, that affect the ranging performance and require tuning for each application. These three parameters are the distance preset, the timing budget and the Region of Interest(ROI). The device has three different distance presets for short, medium and long distance ranging. The second parameter is the timing budget, that basically sets how much time is available for the sensor to complete a range. Bigger timing budget means more accurate measurement, but on the downside it lowers the sampling rate.

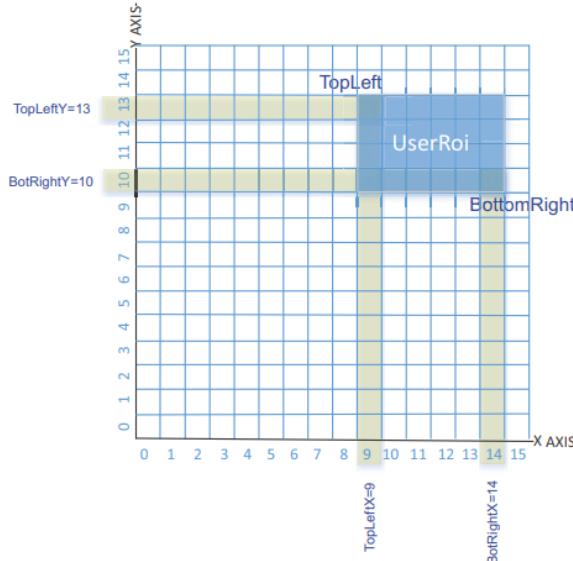
As described in section 2.5 the sensor has a configurable SPAD array that can be used to narrow down the field of view to a specific region. This region in the datasheet is called Region of Interest and will be further referenced to as ROI. By lowering the size of the ROI and changing its position, scanning can be achieved resulting in a higher resolution up to 4x4. Increased resolution means smaller receiving array that lowers sensitivity and maximum distance that can be measured. Scanning also requires more time because each ROI can take time up to the timing budget.

### 3.1.1 Selected operating modes

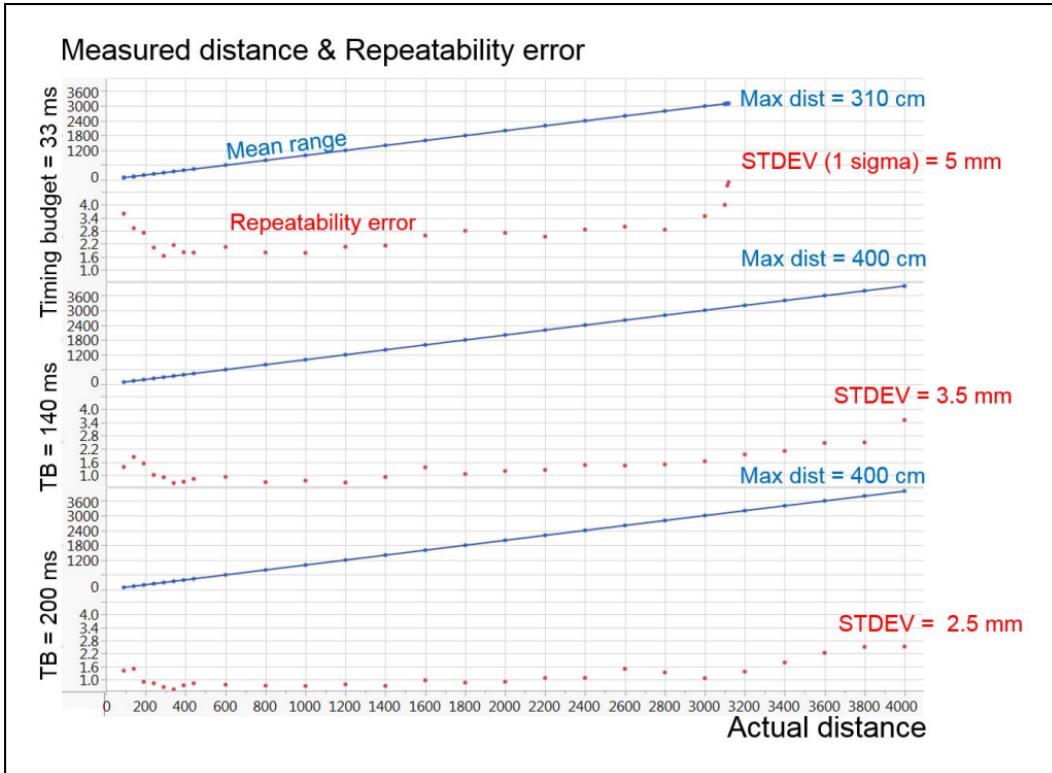
To have a better overview how these parameters affect the ranging performance, a number of operating modes have been chosen for testing. Each operating mode is a group of three parameters that are the distance preset, timing budget and the resolution. The performance of an operating mode is measured by the update rate, maximum ranging distance and standard deviation on a given distance.

According to the VL53L1X datasheet, the 4m maximum distance can only be achieved in long preset, 3m in medium and 1.3m in short. The datasheet contains a plot for maximum distance, standard deviation vs timing budget in long ranging preset with ROI set to maximum size. The timing budgets in the datasheet are 33ms, 140ms and 200ms, therefore I have selected these timings for test measurements.

As an additional setup an operating mode with the highest update rate has been selected, because update rate is expected to improve SLAM performance. This can be achieved by using an inter-measurement period of 20ms and timing budget of 18.5ms, but only in short

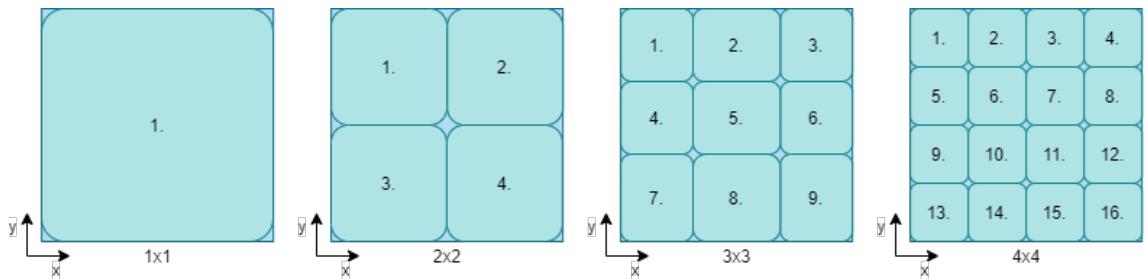


**Figure 3.1:** VL53L1X Region of Interest setting [? ]



**Figure 3.2:** VL53L1X maximum distance, error vs timing budget  
[? ]

range preset and ROI size set to maximum. The estimated highest update rate using this operating mode is 50Hz.



**Figure 3.3:** VL53L1X ROI setups

As seen on figure 3.2 that the sensor provides measurements with a standard deviation under 5mm, but it's not mentioned how the accuracy changes by lowering the ROI size. To investigate the performance of ranging measurements using lower ROI size, four Region of Interest scan patterns have been selected. The regions are selected evenly on the SPAD array to provide 1x1, 2x2, 3x3, 4x4 resolution output. These setups have been selected, because this way the regions do not overlap and are evenly distributed. The sensor would be capable of more complex patterns or even changing the resolution between scans, but these setups should provide a good understanding of the sensor and the effects on the SLAM performance.

In table 3.1 a summary can be found of selected operating modes and a distance to be tested on. These settings are planned to be evaluated incrementally starting with a timing budget of 18.5ms and incrementally moving up to 200ms. With smaller ROI size the

Timing budget	Resolution	Range preset	Distance
18.5 ms	1x1, 2x2, 3x3, 4x4	Short	1 m
33 ms	1x1, 2x2, 3x3, 4x4	Medium, Long	1 m
140 ms	1x1, 2x2, 3x3, 4x4	Medium, Long	2.5 m
140 ms	1x1, 2x2, 3x3, 4x4	Medium, Long	3 m
140 ms	1x1, 2x2, 3x3, 4x4	Long	3.5 m
200 ms	1x1, 2x2, 3x3, 4x4	Long	3.5 m

**Table 3.1:** Selected operating modes

maximum distance is degrading, therefore it is expected that some combinations will not be working properly. For example medium preset on 3 meters is expected to be working only on resolution 1x1, but likely to fail on higher resolutions.

## 3.2 LIDAR layout

As mentioned before, multiple aspects need to be taken into account when placing LIDAR sensors on a quadcopter with the purpose of using the measurements for SLAM. After gaining experience of the performance of each operation mode of the sensor, the next step is to determine the optimal layout and number of sensors to be used.

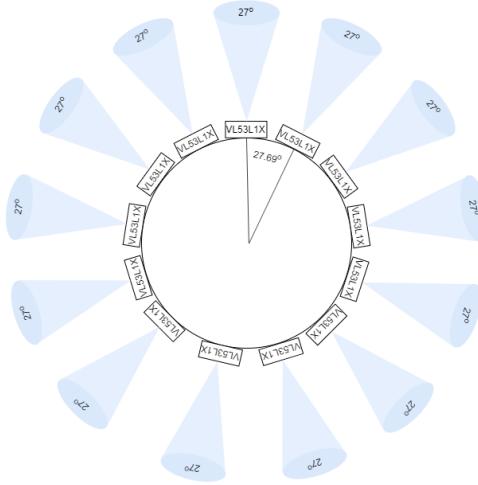
In Cartographer 3D SLAM is basically an extension of their 2D solution, but significantly more complex and harder to tune for optimal performance. With this in mind to better understand the effects of layout it is reasonable to first reduce complexity and test layouts for 2D mapping. Based on the experience gained during this experiment it can be better estimated how many sensors and in what layout are needed for 3D SLAM. A possible outcome is that VL53L1X is not suitable at all for this purpose and a different sensor needs to be used, with more suitable parameters. If mapping and localization accuracy is poor in 2D, it is expected to be worse in 3D.

### 3.2.1 Layout design for 2D SLAM

Without experience with Cartographer or other SLAM algorithms in general, it is not advisable to give an estimation for the number of sensors to be used or the layout of the sensors at this point. My approach is to start with high number of sensors in a reasonable layout, make it work with Cartographer and then incrementally decrease the number of sensors until the result is still acceptable.

For 2D SLAM I chose to place LIDARs evenly to cover the whole  $360^\circ$  range of the horizontal plane, without overlapping fields of the sensors. Due to the high sampling rate and parallel ranging operations, overlapping fields could have a higher probability of crosstalk between neighbors and would produce additional noise in real-world applications, that is better to avoid. Altogether 13 sensors are needed to cover 97.5% of the plane, with  $27.69^\circ$  in between, leaving less than  $1^\circ$  of uncovered zone in between fields. 13 sensors can produce 13 points per scan with ROI resolution set to 1x1 and up to 208 points with the resolution set to 4x4.

In this layout the scanning on the horizontal plane is done purely by the evenly placed sensors around the vertical axis and doesn't take advantage of the movement of the drone during flights, especially turns around the yaw axis. After evaluating the performance of



**Figure 3.4:** Layout of 13 LIDAR sensors

this layout, the number of sensors will be incrementally decreased, by removing a number of sensors at a time. The reduced set of sensors will always be placed evenly, because multicopters can fly in any direction. The flying style is highly dependent on the pilot, so it is best to scan the plane evenly without assuming any dominant flying direction.

The planned bottom limit for the number of sensors is what can be seen on the solution of Bitcraze, the Multi-ranger deck[?] that uses 4 sensors on the horizontal plane facing forward, backwards, left and right. In their demonstration video, the map of a simple room is built in 20 seconds with a ranging frequency of about 10Hz.

### 3.2.2 Layout design for 3D SLAM

The layout for 3D SLAM depends on the experience gained during the evaluation of 2D SLAM, but has more freedom in sensor placement. Because multicopters change their pitch and roll angles to move horizontally, this movement can be used for scanning of the vehicle's environment.

Just like in case of 2D SLAM, in the first layout sensors are placed evenly to cover the whole sphere, without overlapping fields. In the following iterations a number of sensors will be removed until the minimum number of sensors is found that still provide a reliable SLAM performance.

## 3.3 Data collection in Gazebo Simulator

To compare the performance SLAM setups using different layouts and operating modes, I have decided to work on the same set of data in each iteration. ROS offers a great tool to record data that is being transferred in ROS topics into a format that they call rosbag. This tool subscribes to the requested topics and saves all ongoing data into a file with .bag extension. The same tool can be used to play back previously recorded data and the tool will publish messages at the same rate as it was published originally. ROS also provides a C++ and a Python application programming interface for rosbags that can be used to read or write the contents of a bag file.

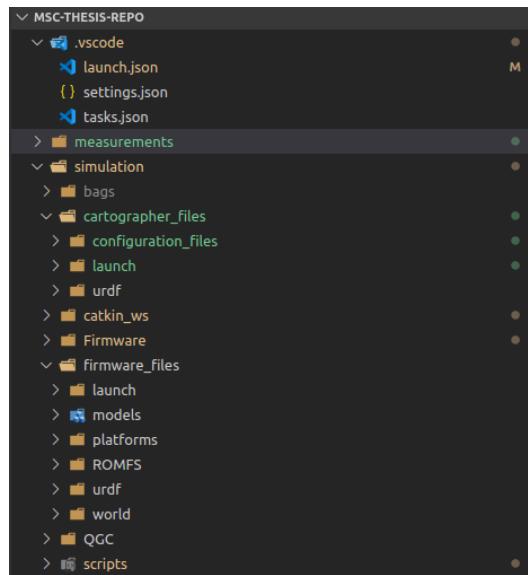
My plan is to place a high resolution 360° LIDAR sensor that covers the whole sphere on a simulated quadcopter, fly it around in a building while recording the necessary topics into a rosbag. The ranges in this rosbag can be then filtered in a way, that its parameters match the selected operation mode and layout. This way the performance of the SLAM using different layouts and sensor operating modes can be compared.

### 3.3.1 Development environment in Visual Studio Code

After building the repository using the guide on PX4 website, I have learned that new models, launch files and world files need to be placed in a subdirectory of the repository that is only generated at build. This also means that if the project is rebuilt, the newly added files might be overwritten. To avoid overwriting on builds, I have started working on these files in a different folder and copy it to their final destination just before starting the simulation.

Cartographer also uses the same mechanism, all files need to be copied to subfolders in its install location. For editing these files I use Visual Studio Code and to make copying easier, I have decided to add a build task in my editor, that copies PX4 files and cartographer files to the corresponding folder. This makes the development much easier and less troublesome.

The project folder structure can be seen on figure 3.5. Measurements folder will contain the output of each SLAM configuration and simulation folder is for ongoing development. Under simulation, the bag folder contains recorded and filtered bag files. All cartographer specific files are in subfolders of cartographer\_files folder. Catkin\_ws folder holds the Cartographer\_ros repository and its dependencies. Firmware is for storing PX4 repository and firmware\_file contains custom files to be copied to the PX4 build directory. QGC stands for QGroundControl, that is not necessary for this simulation, but can be useful in some cases. Lastly scripts folder is for scripts that are used for rosbag filtering, trajectory extraction or other purposes.



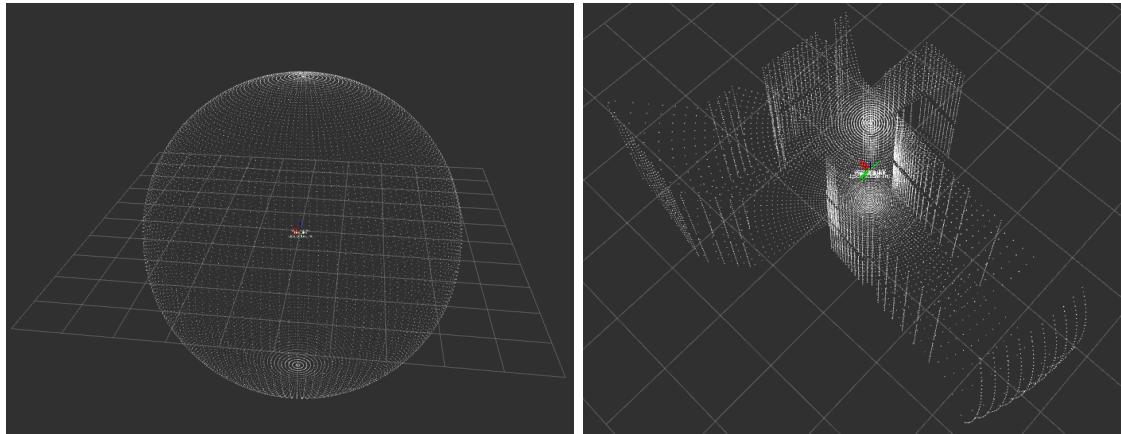
**Figure 3.5:** Project folder structure in Visual Studio Code

### 3.3.2 Adding LIDAR sensor to a quadcopter model

Instead of overwriting files generated during the build procedure of PX4 repository, I have decided to create a custom quadcopter model, that extends the already available Iris drone. Any sensor that is supported Gazebo can be added to it, by editing the model descriptor SDF file.

The equivalent of a LIDAR sensor in Gazebo is called a Ray sensor, that is highly customizable using parameters in the SDF model file. The goal is that the rays of the sensor cover the whole sphere around the quadcopter. The rays of the LIDAR are blocked by the simulated drone so a single sensor cannot be used to cover the whole sphere. I have placed two sensors on the drone, one on the top and one on the bottom as close to the body of the vehicle as possible.

Since the VL53L1X sensors in real-world application would be measuring simultaneously, the simulated LIDARs are also set to scan every angle at once and send all ranges in a single message. The maximum range is set to 4 meters and the update rate is at 50Hz. Gazebo is able to simulate predefined measurement noise, but at first I have decided not to add additional noise. During data filtering these measured point clouds will be down-sampled by averaging on regions and noise can be simulated in addition at this stage. In each scan 128 horizontal and 64 vertical points are sampled per sensor, resulting in a total number of 128\*128 measurement points and angular resolution of 2.8° both vertical and horizontal.



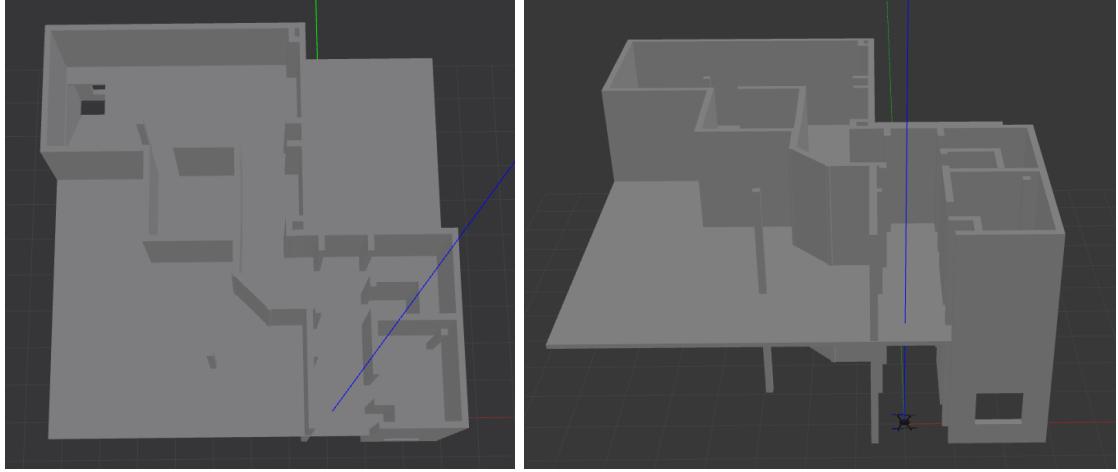
**Figure 3.6:** LIDAR output visualization in rviz

The maximum update rate of each simulated LIDAR is 30Hz, because Ray sensor uses the CPU for computations instead of the GPU. It would be possible to use an alternative called GPU Ray, but I decided not to use it, because of its output message type. Cargorapher expects ranges aggregated into PointCloud2 messages, while Ray sensor's output is PointCloud and the GPU alternative has an output in LaserScan format. PointCloud2 is a newer version of PointCloud and conversion is relatively easy. If 30Hz update rate is proved to be too low during future steps, it is possible to use GPU Ray instead, that requires modifications in the filtering algorithm.

### 3.3.3 Building model

Using Gazebo Building Editor I have created a building model to be mapped. The building is minimalistic and not furnished, but contains enough details like columns and curves to

allow Cartographer to match scans more efficiently. For example SLAM is not expected not work properly on a corridor without any details in range, because it needs a reference points that can be used for scan matching. If each scan looks exactly the same the algorithm will lose track resulting in bad map quality. This is increasingly in focus in this case, because the maximum distance of the VL53L1X sensor to be used is 4m and the sensor has a wide field of view that hides details of faraway objects.



**Figure 3.7:** Two story building in Gazebo

The second floor of the designed building is seen on figure 3.7. The ground level has exactly the same floor map, except it also has a staircase in the top left room, windows and what is important, it has a ceiling. Ceiling is expected to play an important role in 3D slam, adding an extra reference point for vertical positioning. Using this building it is possible in the future to test the efficiency of 3D SLAM by flying up the stairs and around the second floor.

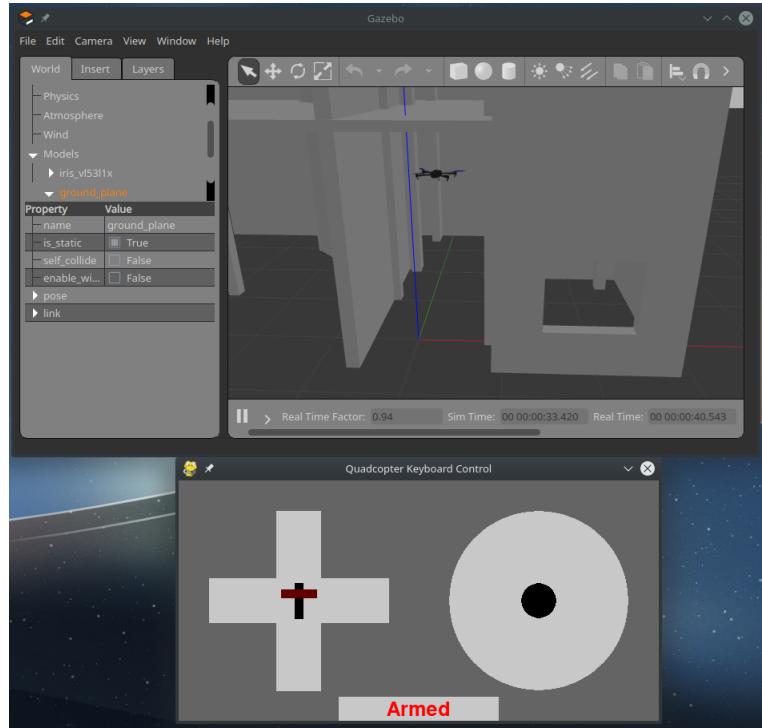
### 3.3.4 Remote controller

After following the previous steps and starting the simulation, the quadcopter with the LIDAR sensor is being spawned at the entrance of the building, but it needs to be controlled to fly around the house. I chose to use a virtual controller that has been developed by a student at the laboratory of my department. The controller uses MAVROS topics to send and receive commands to and from the drone. Using this program, the quadcopter can be flown by using the PC keyboard.

For evaluation of SLAM the recorded flight time is kept low, so the data filtering and Cartographer can run faster. About 2 minutes of data is enough for comparing the performance of SLAM configurations and further flights can take place to further test the evaluated configurations.

## 3.4 Rosbag filtering

Filtering is the process of reading rosbag messages, filtering LIDAR ranges to show behavior of the selected VL53L1X settings and writing the filtered data to a new bag file. Due to the high data rate, rosbags tend to take up lots of disk space. To preserve disk space and increase processing speed, only necessary topics described in 3.5 are being written to the filtered file.



**Figure 3.8:** Virtual remote controller for the simulated drone

The recorded rosbag contains LIDAR measurements on two topics, the top sensor covers the top hemisphere, the bottom one covers the bottom hemisphere. To simulate a VL53L1X sensor behavior, a field of view with a selected opening angle is cut out of the point cloud. Then an average of ranges inside this view is calculated. The averaged distances are then maximized to a set maximum distance threshold and the sampling time is reduced according to the selected device settings. For simulating higher resolutions, the same methodology can be used, by using multiple field of views with smaller opening angles.

The raw measurements read from the collected rosbag are PointCloud type, but Cartographer requires PointCloud2 messages. At the end, the filtered points are converted to PointCloud2 and written to the filtered bag file.

### 3.4.1 Update rate

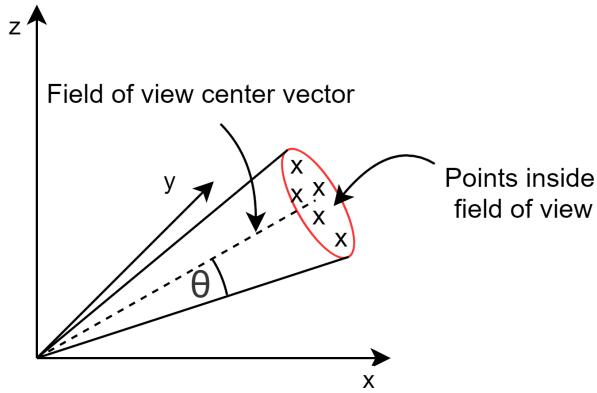
LIDAR messages are recorded in a rosbag at a certain rate and it needs to be down-sampled to match the desired update rate. Every PointCloud message has a header field that contains the current time at publish. During filtering each message is read from an input bag in order as they were published. A PointCloud message is only considered for filtering, if the time difference between the last written and the current message is greater than the defined update rate. Otherwise the message is discarded.

Using this procedure the update rate will not match exactly the desired rate, but keeps data integrity and order with other messages.

### 3.4.2 Field of View

Each PointCloud message contains a complete scan from either the top or bottom sensor, that scan the top or bottom hemisphere accordingly. As the message type suggests, these ranges are represented by point clouds, a list of 3 dimensional point coordinates in reference to the sensor coordinate system. During angle filtering this point cloud is being down-sampled to lower resolution that simulates the sensors placed on the quadcopter.

VL53L1X sensor has a cone shaped field of view in which the reflected signal is detected and converted to a range measurement. To match this behavior, a cone shaped subset of the points are selected for further processing. A point is selected if the angle between the point and the center of the field of view is less than a predefined  $\theta$  angle threshold.



**Figure 3.9:** Point cloud angle filter

The angle between two vectors, or in this case between a point from the point cloud and the center vector can be calculated using the dot product of the two. Equation 3.1 and 3.2 are two forms of dot product. These are used to derive equation 3.3 that is used to calculate the angle. In this case vectors  $\vec{u}$  and  $\vec{v}$  are the center vector and a selected point from the point cloud.

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos(\alpha) \quad (3.1)$$

$$\vec{u} \cdot \vec{v} = u_x v_x + u_y v_y + u_z v_z \quad (3.2)$$

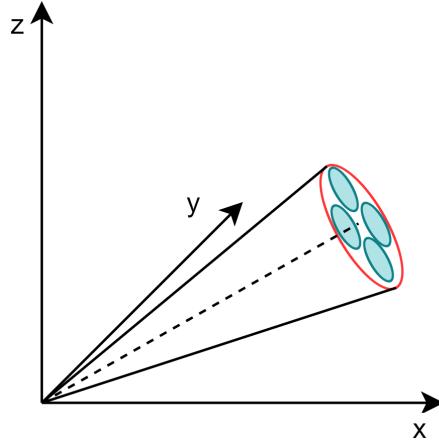
$$\alpha = \cos^{-1} \frac{u_x v_x + u_y v_y + u_z v_z}{\|\vec{u}\| \|\vec{v}\|} \quad (3.3)$$

$$\|\vec{u}\| = \sqrt{u_x^2 + u_y^2 + u_z^2}, \|\vec{v}\| = \sqrt{v_x^2 + v_y^2 + v_z^2} \quad (3.4)$$

For every point inside the point cloud this angle is calculated and points that have an angle less than the predefined  $\theta$  are added to the list of points inside the field of view. As a final step average of the selected points is calculated. Points once again are represented by coordinates, therefore average is calculated by averaging the x, y and z coordinates resulting in the coordinates of a single point.

The steps described above are for the filtering a single sensor, but the same steps apply for multiple sensors. Higher resolutions are constructed by multiple field of views, with a

lower angle threshold of  $\theta$ . Figure 3.10 shows an example for field of view selection for a resolution of 2x2. Field of view center line and angle threshold are not drawn, to keep the figure simple.

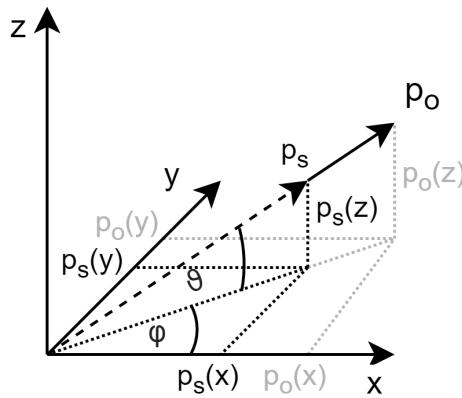


**Figure 3.10:** Representation of 2x2 LIDAR resolution filtering

Points inside a PointCloud message always take the same order, therefore it's enough to filter for angles only once and use the indexes of these points from that on. This way lots of processing capacity can be saved and it proved to be much more efficient on high number of simulated sensors.

### 3.4.3 Distance filter

Angle filter returns a single point for each predefined field of view. For each sensor it can mean a single point or up to 16 points in case of a resolution of 4x4. As the second step, distance filter receives these points and calculates the distance of each point from the origin, basically the same as calculation of length of a vector. If the distance is greater than the defined maximum distance, the length of the vector is shortened to the maximum distance. On figure 3.11  $p_o$  represents the point and  $p_s$  is the shortened vector.



**Figure 3.11:** Point cloud distance filter

$$\|\vec{p}_o\| = \sqrt{p_o^x + p_o^y + p_o^z} \quad (3.5)$$

To shorten the length of a vector, all three coordinates need to be recalculated in such way, that the attitude of the vector does not change. To do so, vertical  $\vartheta$  and horizontal  $\varphi$  angles for vector  $p_o$  are calculated using equation 3.6 and 3.7.

$$\vartheta = \tan^{-1} \frac{p_o^z}{\sqrt{p_o^{x^2} + p_o^{y^2}}} \quad (3.6)$$

$$\varphi = \tan^{-1} \frac{p_o^y}{p_o^x} \quad (3.7)$$

Using the vertical and horizontal angles, the coordinates of the shortened  $p_s$  is calculated using trigonometric equations 3.8, 3.9 and 3.10.

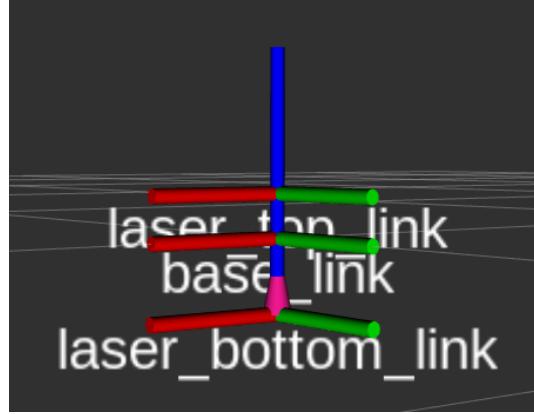
$$p_s^x = d_{max} \cdot \cos \vartheta \cdot \cos \varphi \quad (3.8)$$

$$p_s^y = d_{max} \cdot \cos \vartheta \cdot \sin \varphi \quad (3.9)$$

$$p_s^z = d_{max} \cdot \sin \vartheta \quad (3.10)$$

### 3.5 Cartographer environment

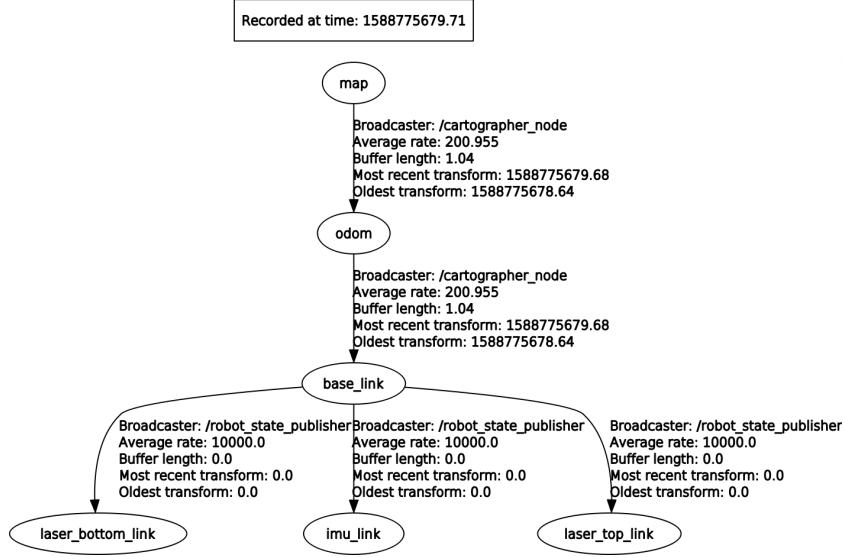
Cartographer can be installed on Ubuntu operating system easily by using the apt-get tool, but this method would overwrite the system installed Protobuf with a newer version. Protobuf is used by Gazebo and it requires a certain version, replacing it with another version causes incompatibility and simulations cannot be started. For this reason I have installed Cartographer and Cartographer\_ros packages inside a Catkin workspace and set them to use locally installed Protobuf inside the same workspace. Using these settings, Cartographer and Gazebo can be used on the same PC.



**Figure 3.12:** Cartographer top, bottom LIDAR and IMU link relations

To attach the LIDAR sensors onto the simulated quadcopter in Gazebo, two links have been introduced that connect to the base\_link of the drone, one on the top and one on the bottom. The LIDAR scanner messages are sent in reference to these links and therefore Cartographer needs to know the transformation from the sensor frames to the base\_link frame. Two official packages are required called robot\_state\_publisher and

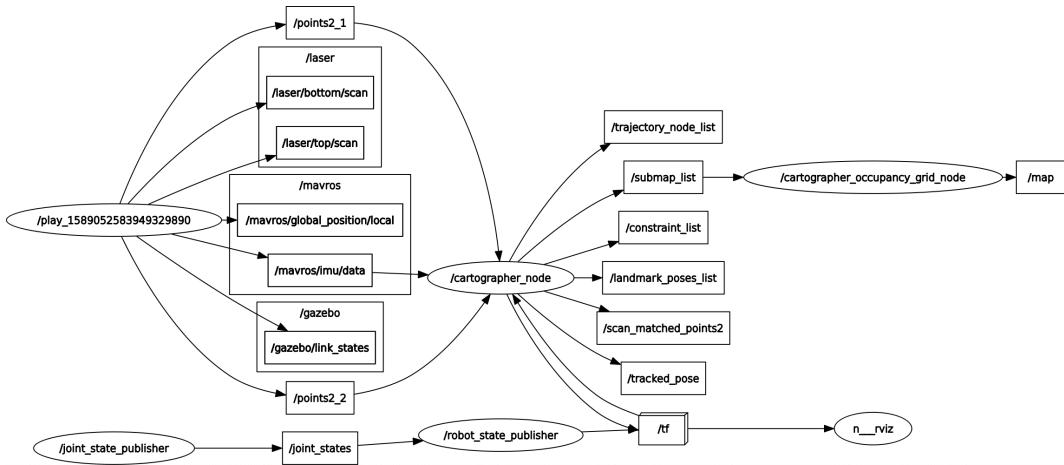
joint\_state\_publisher. These take a robot\_description parameter that can be set using an urdf file and they start publishing the link and joint relations. Besides the LIDAR holder links, a link for IMU transformation is also need to be added. The simulated IMU in the PX4 repository is used and it is located in the origin of base\_link. Laser\_top\_link is located 6cm up, laser\_bottom\_link is located 10cm down on the z axis in reference to base\_link as seen on 3.12.



**Figure 3.13:** Cartographer tf tree

While running Cartographer on filtered data, the following tf tree is built as on 3.13. Transformations between map, odom and base\_link are published by Cartographer and the rest is published from the created urdf file. This tf tree matches the recommendation from the Cartographer documentation website [? ].

IMU input for 3D mapping is a must and optional for 2D mapping. Cartographer waits for IMU to be supplied on topic /imu. PX4 simulation publishes IMU data to /mavros/imu/data, but this topic can be remapped to /imu using the launch file without the need of any extra nodes. Filtered range messages with PointCloud2 type can be accessed on topics /points2\_1 and /points2\_2. List of nodes and interactions via topics can be seen on figure.



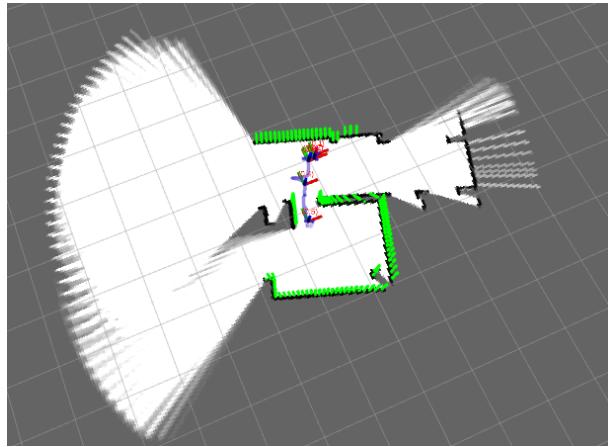
**Figure 3.14:** Cartographer nodes and topics

### 3.5.1 Configuration and tuning

Cartographer can be configured using a lua configuration file. Multiple range measurement types are supported, but the filtered rosbag contains PointCloud2 messages on two topics. The num\_point\_clouds parameter is set to 2 to match the two topics coming from the top and bottom sensors.

Cartographer applies a bandpass filter on the incoming range messages to filter out irrelevant measurements. In every case max\_range parameter needs to be set a bit lower, than the maximum distance parameter used for filtering. This is important, because the algorithm inserts a range with a distance under max\_range as a hit, but inserts anything above as a miss. Wrongly setting this parameter results in adding all range measurements as hits and no misses will be inserted that will confuse cartographer.

Each sensor setup require the SLAM algorithm to be tuned and therefore a new configuration file is created for every setup being tested. Tuning starts with setting POSE\_GRAPH.optimize\_every\_n\_nodes to 0 to disable global SLAM. Good and clear submaps are needed for global SLAM to find loop closures. Submaps are the building blocks of maps. On small maps like the building I chose for LIDAR setup evaluation, global SLAM is not making much difference, but when tuning submaps it can cause unwanted artifacts that are hard to differentiate from submap errors.



**Figure 3.15:** Building of submaps visualized in rviz

Typically a small number of parameters need to be tuned for submaps to be working. These are the following:

- TRAJECTORY\_BUILDER\_nD.ceres\_scan\_matcher.translation\_weight
- TRAJECTORY\_BUILDER\_nD.ceres\_scan\_matcher.rotation\_weight
- TRAJECTORY\_BUILDER\_nD.ceres\_scan\_matcher.occupied\_space\_weight
- submaps.num\_range\_data

The first parameter sets how high the cost is for cartographer to move the a scan. In my experience, it needs to be set low when used on a drone, because it changes place quickly. The second parameter sets how much trust should be put on the IMU data. The lower this parameter is set, the easier it is to rotate a scan and be matched against other scans in the submap. Occupied space weight is sets the trust in the LIDAR sensor. I haven't found this parameter particularly useful. Changing the submap size, by num\_range data

parameter is important, because by changing the update rate, different number of range measurements are needed, while the submap scope should not change.

### 3.5.2 Trajectory extraction

Evaluating the mapping quality is complex relative to evaluating the location tracking performance of a SLAM setup. Gazebo provides a topic where all link states are published and this transformation can be used as ground truth and trajectory extracted from SLAM can be compared to it.

I found two ways to extract trajectory from Cartographer. One is to create a node that subscribes to `/tf` topic and saves map to `base_link` transformations that is published by Cartographer. The other way is to save the internal variables of the algorithm into a `pbstream` file and use a script to read the pose estimations from there. For some reason published transformations proved unreliable on my tests, but the published trajectory node list on topic `/trajectory_node_list` contains the correct pose estimates.

Cartographer can run in mapping mode and pure localization mode. In localization mode, it only stores variables for several seconds and for this reason the `pbstream` file does not contain the whole trajectory. In this mode, trajectory can only be extracted by saving pose from `/trajectory_node_list` topic. For mapping mode I have decided to use trajectory extracted from the `pbstream` file.

## 3.6 SLAM performance evaluation

As the name SLAM suggests and described before, it focuses not only on mapping but simultaneous localization too. On the other hand evaluation of mapping and localization can be done separately. This thesis mainly focuses on localization, while mapping is secondary. For this reason only localization accuracy is being measured quantitatively and mapping accuracy is only evaluated by visual inspection.

By SLAM performance I mean the accuracy of the localization of each setup. While the number of sensors used and their placement are important parameters of a setup, these are not included in the evaluation process. The goal of evaluation is to give an objective measure of the localization accuracy, independently from the hardware implementation.

In order to compare SLAM setups, an objective measure of goodness need to be introduced. Based on this quantitative measure or score, the setups can be compared. Localization is chosen to be the main focus and it can be measured by comparing the trajectory output of the algorithm to a ground truth. As described in 3.5.2 the list of estimated poses can be extracted from the `pbstream` file, provided by Cartographer. On topic `/gazebo/link_states` transformations are available for `base_link` that were used for visualization of the quadcopter in Gazebo. This data is the most accurate position information available and therefore it is chosen to be used as ground truth for the drone's trajectory.

Cartographer can either be run in normal operation or in pure localization mode. Pure localization mode means, that the algorithm only keeps a small number of submaps in the memory at a time and matches them against submaps of a previously built map. Further building of the map is disabled in this mode and submaps will be discarded once they become obsolete.

In case of Cartographer, pure localization mode is basically a stripped down version of the normal operation mode and for this reason building good quality submaps is also

essential for the algorithm to be working accurately. Tuning procedure can happen in normal operation mode and then used in localization mode afterwards. A setup that is able to produce good quality submaps, means it will also work accurately in localization mode, therefore I chose to evaluate and compare setups in normal operation mode.

### 3.6.1 Down-sampling

Link states are much more often published by gazebo, than the extracted trajectory by Cartographer. To compare the two sets of data, they need to have the same size. To achieve that, ground truth data is down-sampled to the size of the SLAM trajectory.

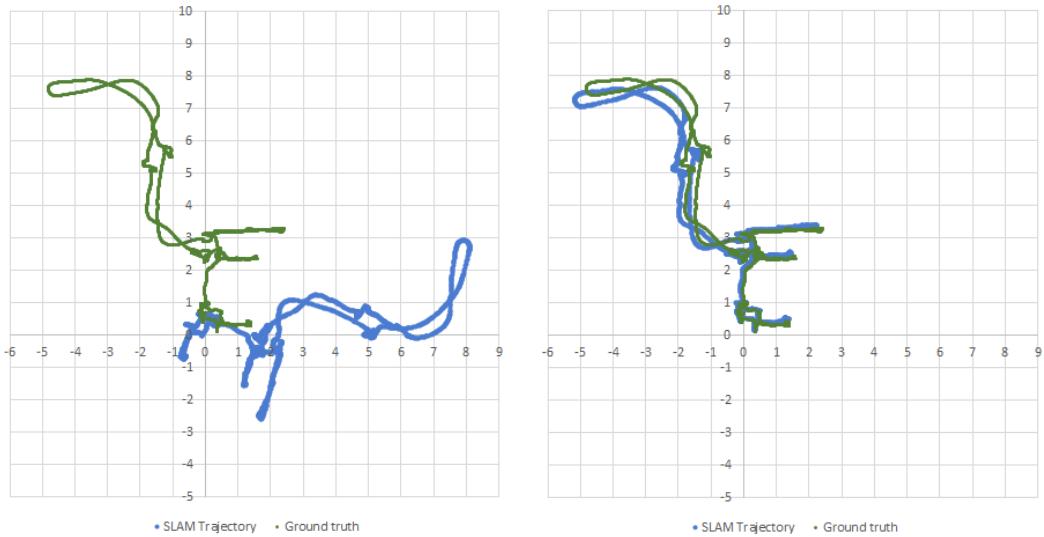
### 3.6.2 Offset compensation

Before error calculations, the SLAM trajectory needs to be rotated and moved to align with the ground truth data. The distance and rotation offsets are calculated using the first pose from both datasets. The distance offset equals to the difference of the x,y and z coordinates of the poses, while the rotation offset is the difference between the yaw angles. In my initial tests there were no roll and pitch offsets present, because of the use of IMU of Cartographer.

To compensate for different starting points, the calculated offsets on each axis needs to be subtracted from every data point present in SLAM trajectory. For the rotation offset, the following trigonometric formulas are used, where  $\theta$  is the offset on yaw angle.

$$x' = x \cdot \cos\theta + y \cdot \sin\theta \quad (3.11)$$

$$y' = -x \cdot \sin\theta + y \cdot \cos\theta \quad (3.12)$$



**Figure 3.16:** Extracted trajectory against ground truth before and after offset compensation

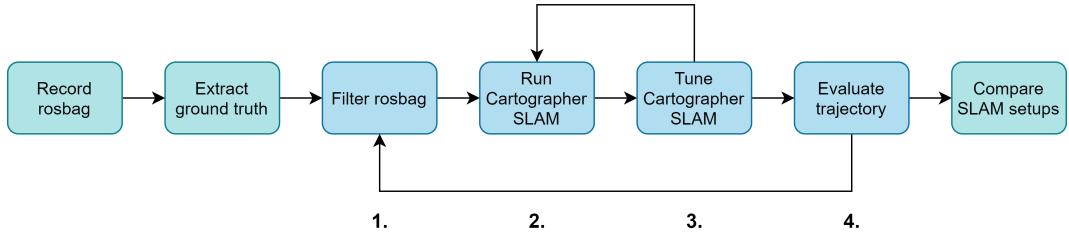
### 3.6.3 Trajectory error calculation

As a quantitative measure I chose to use the error of the estimated trajectory, calculated using the RMS error formula 3.13. The RMS error is calculated on all three axes,  $\hat{c}_i$  is one coordinate of the estimated pose and  $c_i$  is the according coordinate of the pose from ground truth trajectory.

$$RMSError(c) = \sqrt{\frac{\sum_{i=1}^n (\hat{c}_i - c_i)^2}{n}} \quad (3.13)$$

After calculating RMS error on all three axes, to get the cumulated trajectory error the values are summed. The result is a single positive number in meters, that can be used to compare the accuracy of different SLAM setups.

## 3.7 Complete workflow



**Figure 3.17:** Complete workflow

The complete workflow is as seen on figure 3.17. As preliminary steps a rosbag is recorded and ground truth trajectory is extracted from it. After this step, all SLAM setups will be evaluated based on this data. To evaluate a selected setup, first the rosbag is filtered according the layout and sensor settings. Then Cartographer SLAM is run and tuned for this dataset. At step 4 the RMS error is calculated based on the estimated trajectory from the SLAM algorithm.

These 4 steps are done for every selected SLAM setups and as a final step the results will be compared.

### 3.8 Plans for hardware implementation

On a real-life drone, the communication of LIDAR sensors and measurement forwarding to ROS topics need to be solved. There needs to be a device that communicates with the VL53L1X LIDAR sensors using I2C protocol and forwards messages to a ground station that handles data collection and processing.

The data collector device needs to be fast enough to keep up with the pace of 10-20 sensors, each with a maximum of 50Hz update rate. The distance data from VL53L1X sensor consists of 12bytes, so by designing for 20 sensors with 50Hz update rate generate about 12kb per second data rate. This load is calculated using the effective data size, but on the I2C bus other configuration messages need to be sent, which means even higher load.

To evaluate the SLAM algorithms on real-world measurements, data needs to be collected. For evaluation it is not necessary to send data in real-time on wireless network, but saving data on an SD card serves this purpose. Logging on SD card has low complexity and high speed.

PX4 is an open-source firmware, it seems an easy choice to use the Pixhawk 4 autopilot board as a data collector device. It has an SD card slot for logging and I2C connectors available to communicate with the LIDAR sensors and even comes with an I2C splitter board to connect more sensors. The microcontroller inside Pixhawk 4 is powerful, it has an ARM Cortex-M7 core running on 216MHz. It is used to run highly timing sensitive control loops to produce actuator values for stable flights, besides many other tasks. If these control loops are delayed by other processes, that can cause instable flight or even crash.

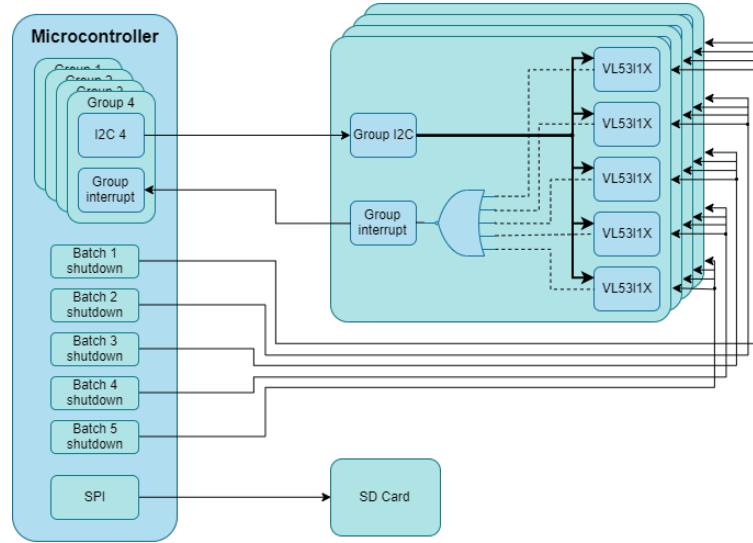
Because of lack of deep understanding of PX4 firmware, to make sure that timing of control loops are untouched, I decided to design a standalone data collector, using a dedicated microcontroller that handles I2C communication with the LIDAR sensors and writes data to an SD card using SPI protocol.

#### 3.8.1 Standalone data collector design

In I2C protocol every device needs to have a 7 bit address, that needs to be unique on every I2C line. VL53L1X sensor has a default address of 0x29 when the sensor is booted, but it can be changed by using I2C commands. Address conflicts need to be avoided by having exactly 1 sensor active per channel. It is hard to find a microcontroller that has 20 I2C buses, one for each sensor, so multiple sensors need to be placed on the same bus. By having multiple sensors on the same channel, the sensors need to be released from reset one by one to change their addresses, this way address conflicts can be avoided. VL53L1X can be kept in reset by pulling its shutdown pin to ground and activated by bringing it to supply voltage.

The sensors have altogether 4 lines that need to be driven: 2 I2C, 1 shutdown and 1 interrupt pin. In the simplest case, by connecting each wire to the microcontroller, 80 wires would be needed for 20 devices and would take 42 pins of the microcontroller. Although it is possible to find a microcontroller that has enough pins and connect all sensors one by one to the same I2C bus, but by grouping them into groups of 5 and connecting each group to different I2C buses, a significant improvement can be achieved on the number of connections.

Sensors on different I2C buses can have the same address, therefore the same 5 addresses can be used in each group. Shutdown pin of sensors sharing the same addresses can be common, this way number of wires needed for shutdown is reduced from 20 to 5.



**Figure 3.18:** Design of data collector

Handling 20 interrupt lines can be overwhelming for any CPU, while bringing little or no extra benefit. Using only one interrupt line per group is enough to signal the microcontroller that measurements are ready to be read out in all sensors in that group. Interrupt pins are active-low, so to produce the group interrupt signal a NOR gate needs to be used. When the state of all interrupt pins are low, group interrupt signal will become logical 1 and 0 otherwise.

In the simplified design seen on figure 3.18 only 17 wires are connected to the microcontroller, while no functionality is lost. This solution also makes software development easier and the built system is clearer.

# Chapter 4

## Implementation

In this chapter the implementation steps are introduced of the system described in 3. Two LIDAR sensors are placed on a simulated quadcopter in Gazebo, that cover the top and bottom hemispheres around the drone. Using the data collected in the simulator, these range measurements are filtered to simulate a number of VL53L1X sensors with a predefined operating modes and orientations.

In order to simulate the VL53L1X sensors accurately a set of parameters, capabilities and limitations of the LIDAR sensor is measured. Data is collected using selected operating modes on selected distances that are described in 3.1. Average distance, standard deviation, sampling rate and typical range status are extracted.

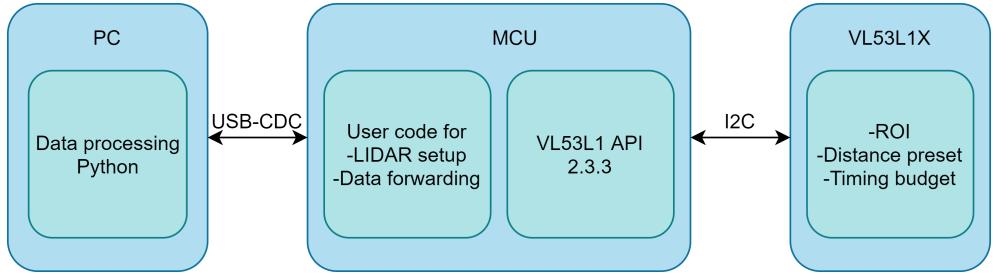
The two main components in the evaluation of VL53L1X sensor are a microcontroller and a PC. The microcontroller communicates with the sensor and sends the raw measurements to the PC that runs data processing and evaluates the received data. I chose to separate data collection and processing, because data processing is significantly easier in high level programming languages, like Python and is much easier to adjust. It also enables data collection and offline processing.

### 4.1 VL53L1X measurement setup

For each measurement a combination of parameters need to be configured on VL53L1X, these are the timing budget, resolution and distance preset. The configuration is done via I2C protocol by the microcontroller. VL53L1 API from STMicroelectronics is used for interfacing the sensor with the most recent version at the time: 2.3.3. After completing boot the recommended initialization steps are used, as described in example codes and the API user manual [? ].

Each measurement is triggered in single shot mode, that means exactly one ranging is done by the sensor and afterwards it waits until a new command arrives. Single shot is preferred over continuous measurements, because if the resolution is set to higher than 1x1, measurements need to be done on multiple ROIs and ROI has to be configured before triggering each measurement. In continuous mode the ROI cannot be modified between measurements. A scan is complete if ranges are done on all ROIs of interest. The completed scan is then forwarded for processing to the PC via USB-CDC protocol.

I chose Python as a programming language to read data from USB-CDC protocol and process the incoming data, because Python is rich in data processing tools and enables quick development. The script first reads 50 scans coming from the MCU, then closes



**Figure 4.1:** VL53L1X measurement workflow

communication and works on the collected data. The features extracted area the average distance, standard deviation, typical range status and sampling time. Average and standard deviation are calculated based on the whole population as in equations 4.1 and 4.2, where  $s$  and  $r$  are the indices of scan and range,  $S$  and  $R$  are the number of scans and ranges accordingly.

$$d_{avg} = \frac{1}{S \cdot R} \sum_{s=0}^S \sum_{r=0}^R d_{sr} \quad (4.1)$$

$$d_{std} = \sqrt{\frac{1}{S \cdot R} \sum_{s=0}^S \sum_{r=0}^R (d_{sr} - d_{avg})^2} \quad (4.2)$$

#### 4.1.1 Comparison of selected operating modes

VL53L1X LIDAR sensor has an adjustable timing budget from 18.5ms all the way to 1 second, that determines the maximum time a ranging operation can take. To see the effect of timing budget with different ROI setups and distance modes, 4 timing budgets have been selected.

With 18.5ms budget the sensor operates on the highest possible measurement frequency of 50Hz. This can only be achieved in short distance mode without the possibility of modifying the ROI. The second selected budget is 33ms, that is the lowest possible value to be used with long distance preset. 140ms is the lowest value that can provide measurements up to 4 meters. Lastly 200ms budget is selected to see how a higher timing budget affects accuracy. It is also the value used in the VL53L1X datasheet[?] to demonstrate ranging accuracy.

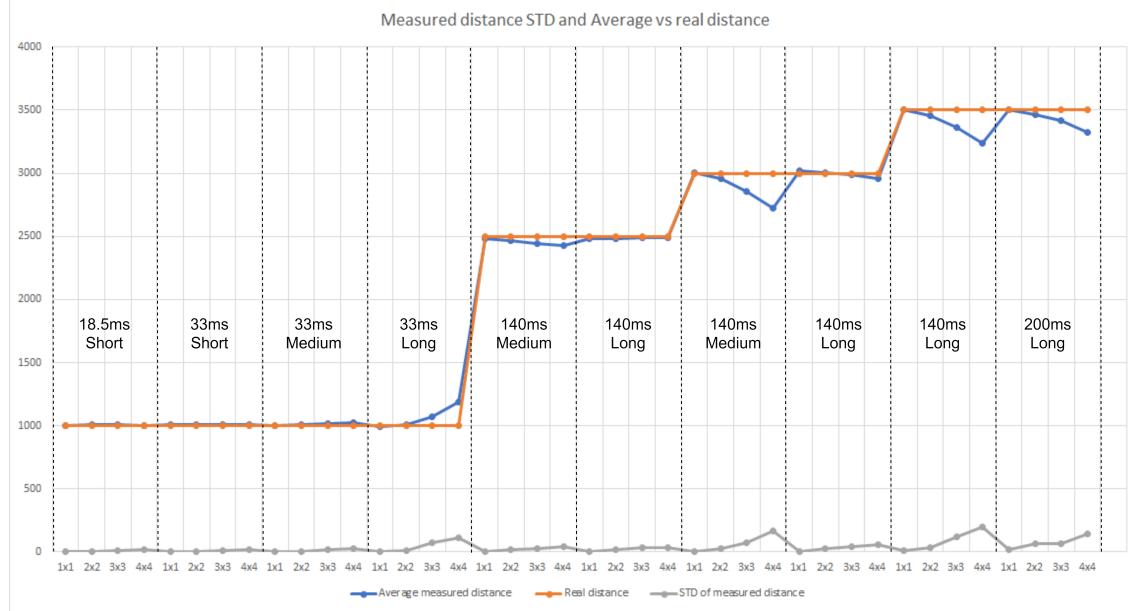
The calculated average distance and standard deviation against the real distance can be seen on figure 4.2, std separate is on 4.4. The measured sampling time accordingly is visible on figure 4.3.

Based on the completed measurements, it can be seen that sampling time is independent from the distance preset in use and is directly proportional to the number of ROIs in a scan and the timing budget. 50Hz sampling rate is not achieved even with 18.5ms timing budget, because the LIDAR was used in single shot mode and each measurement needs to be triggered by the MCU. Sampling time can be increased for setups with 1x1 resolutions by switching to continuous measurement mode.

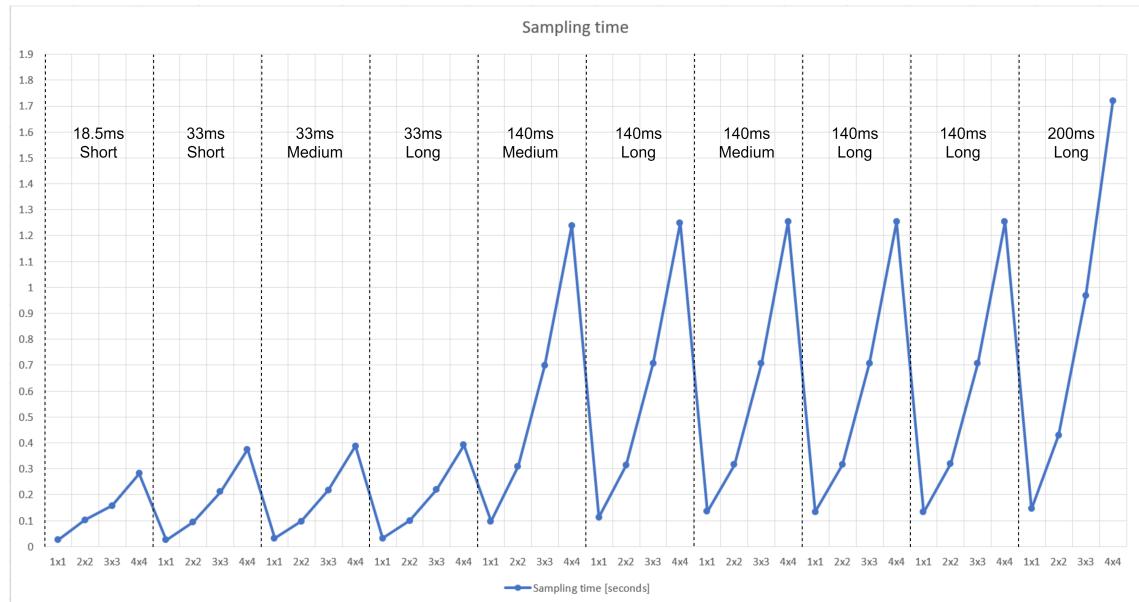
In every setup on each distance the most accurate measurement is always with resolution 1x1. By increasing the resolution, therefore lowering the receiver SPAD array size, the

measured distance starts to deviate from the real distance and STD increases. This can be seen on 2.5m distance with medium preset. Long distance preset however produces more accurate measurements even on 3m, so it is preferred over medium preset.

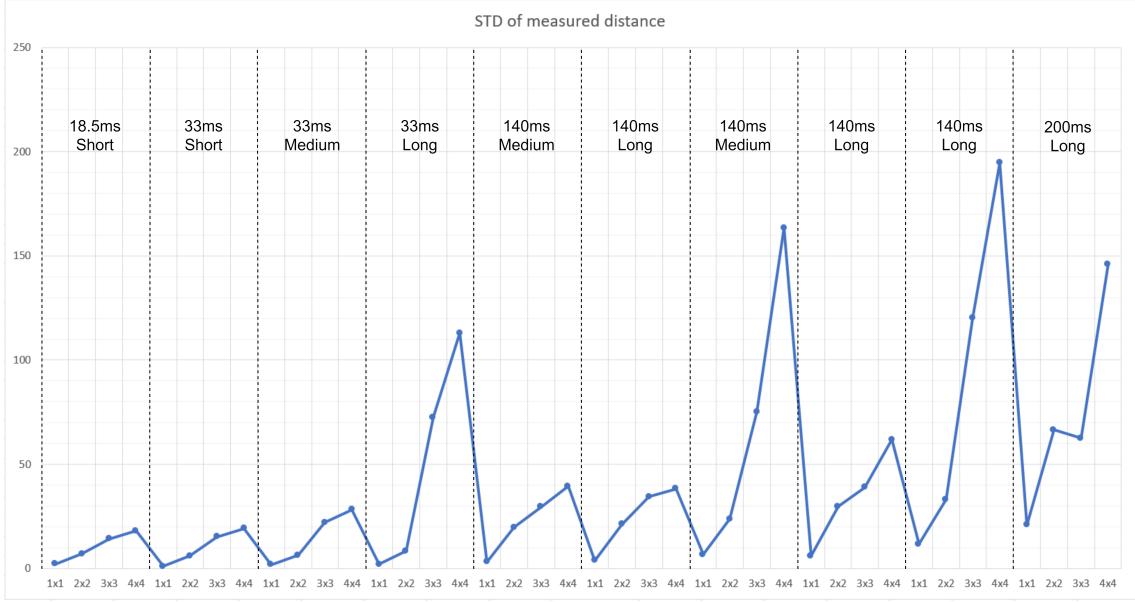
Measurements with long distance preset and 140ms timing budget follows the real distances even on 3 meters, but starts to deviate on 3.5 meters with 2x2 resolution and above. Raising the timing budget to 200ms reduces the error and standard deviation, but doesn't eliminate completely.



**Figure 4.2:** VL53L1X measured distance STD and Average against real distance



**Figure 4.3:** VL53L1X range sampling time in different operation modes



**Figure 4.4:** VL53L1X range standard deviation in different operation modes

#### 4.1.2 Detailed distance measurements

Based on the measurements described in the previous section, I chose a small set of LIDAR settings to further investigate and see how these behave on different distances. These settings are candidates to be used in simulations. The goal with the setting selection is to choose the ones that may have greater affect the SLAM performance.

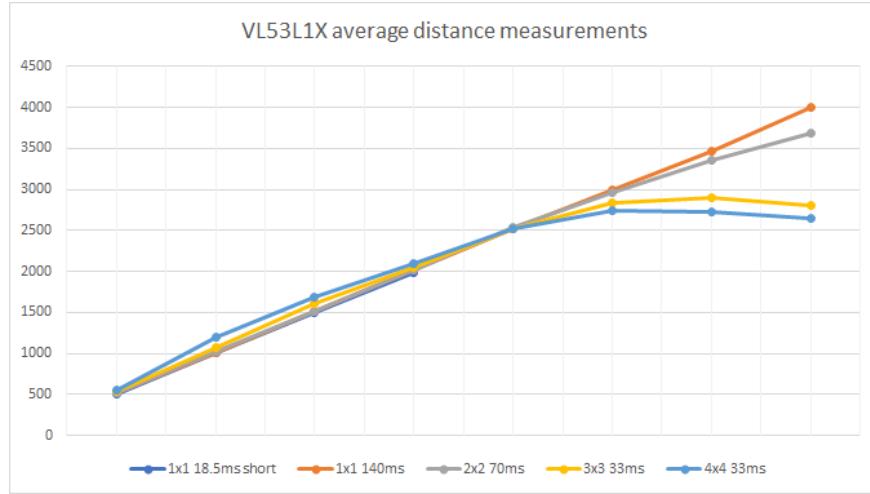
After the initial tests with Cartographer SLAM, it's safe to assume that sampling time shouldn't be greater than 500ms. This is not a strict threshold, but with a higher sampling time, the algorithm would have a hard time to follow the position of the drone and insert new scans precisely. The second assumption is that more points per scan increase the performance of scan insertion, because more points create a more detailed and therefore more unique image.

I chose 5 setups for further investigation, 4 with each resolution and the fastest one with 18.5ms timing budget. The lowest timing budget to be used with 4x4 resolution is 33ms, that is expected to be just under the previously described threshold of 500ms per scan. The same budget is chosen for 3x3 resolution, while 2x2 has lower number of ranges per scan so a higher budget, 70ms is selected to have more accurate measurements. As for 1x1 18.5ms and 140ms timing budgets are selected to see the difference between a faster and a slower setup.

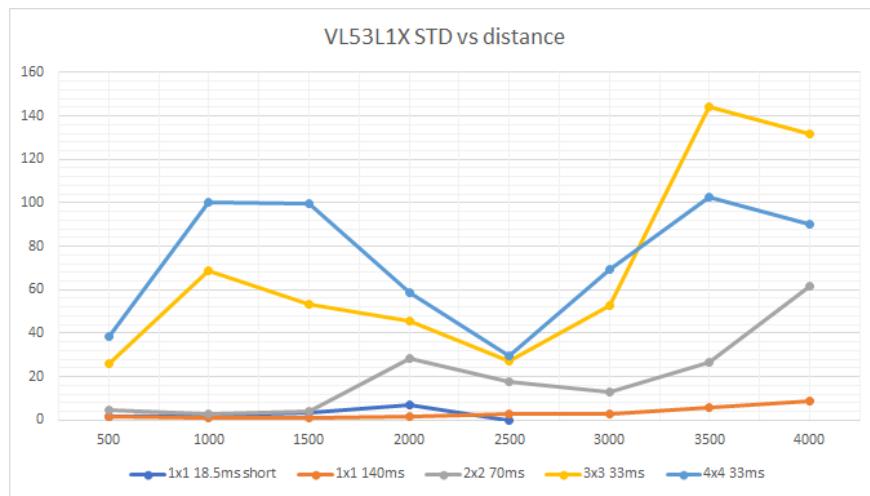
Each setup has been measured on distances starting from 0.5m to 4m with a step size of 0.5m. The measurements can be seen on figures 4.5, 4.6, 4.7. In the end all setups turned out to have a lower sampling time than 500ms. It is interesting, that sampling time turned out to be higher on smaller distances and lower on high distances. This might be because the internal algorithm of VL53L1X sensor does longer measurements to reassure that these ranges are not artifacts coming from reflections or crosstalk.

As expected the setup with 1x1 resolution and 140ms timing budget can range the farthest with only 9mm standard deviation. The setup with 2x2 resolution deviates from the real

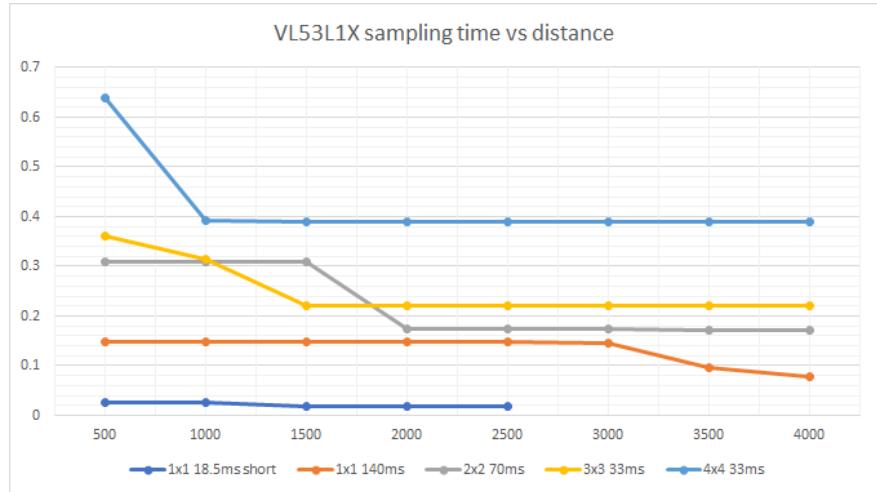
distance by 16cm on 3.5m, but is still accurate on 3m. The setups with 3x3 and 4x4 resolutions can measure up to 2.5m, but deviate on farther measurements.



**Figure 4.5:** Detailed distance measurements [mm]



**Figure 4.6:** Standard deviation of selected measurements [mm]



**Figure 4.7:** Sampling time of selected measurements [s]

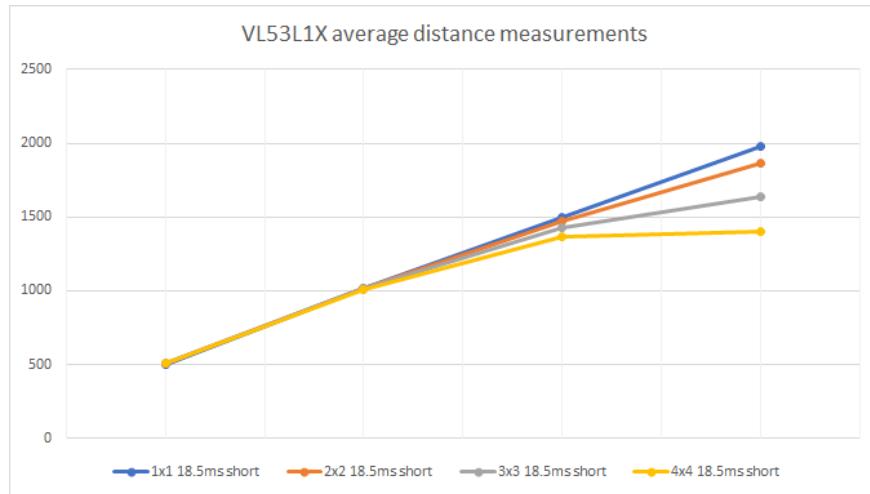
#### 4.1.3 Detailed measurements with the fastest setup

18.5ms timing budget with short distance preset can produce the fastest ranging measurements up to 1.6m according to the datasheet. The short preset also has the best ambient light immunity, a favorable quality. Due to these attributes I have decided to do more measurements using these settings in combination with all resolutions.

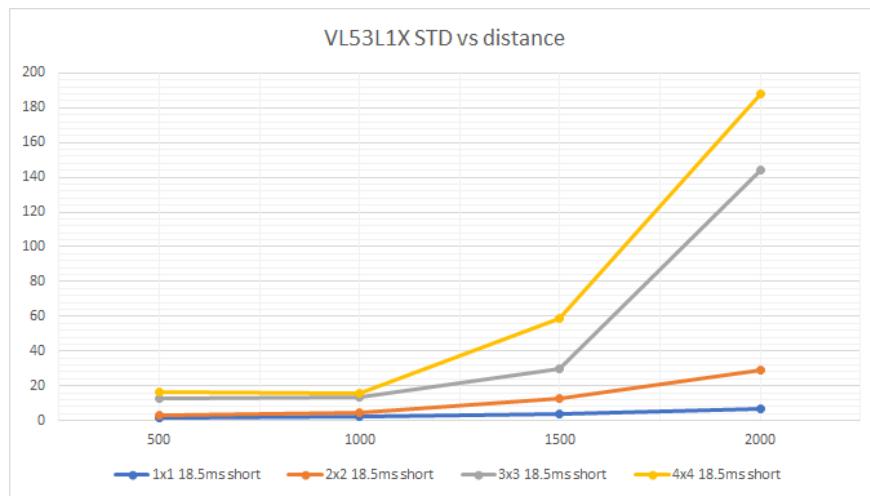
Surprisingly the sensor with these settings was able to measure up to 2m with only 7mm standard deviation. Above 2m no ranges were successful, the sensor always reported 0 distance instead.

In comparison to the previous results the sampling time for each resolution has decreased significantly. In case of 4x4 resolution, the sampling time has decreased from 390ms to 265ms and the maximum distance from 2.5m to about 1m. The increase in update rate might be beneficial for the SLAM algorithm.

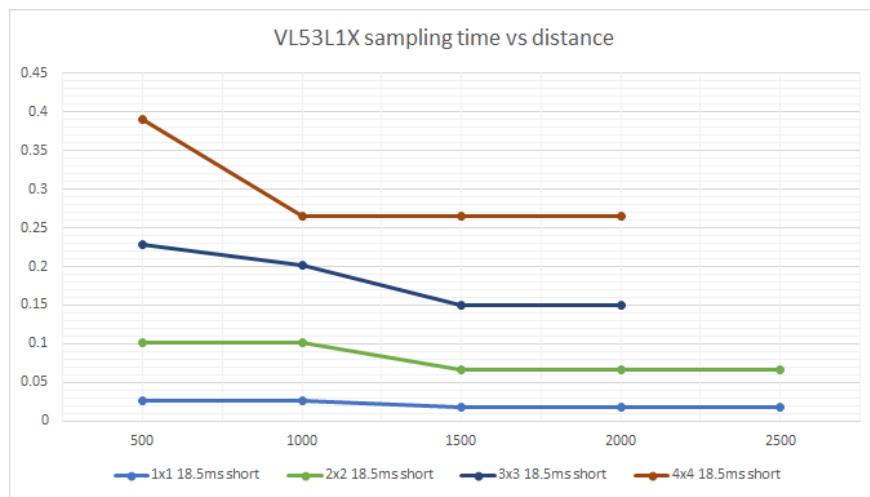
Surprisingly the update rate for 1x1 resolution reached 56Hz, that is higher than the maximum update rate according to the datasheet. The reason for the increased update rate is that 50Hz was calculated using continuous measurement mode that places a 1.5ms delay between ranges. In this scenario the device was used in single-shot mode and there was no delay between measurements. Although this method puts a high load on the host microcontroller that is not favorable in most applications.



**Figure 4.8:** Detailed distance measurements [mm]



**Figure 4.9:** Standard deviation of selected measurements [mm]



**Figure 4.10:** Sampling time of selected measurements [s]

## 4.2 2D SLAM evaluation

During the evaluation both in 2D and 3D the same collected data will be used. The length of the collected rosbag is 144 seconds and contains all messages coming from all nodes on all topics. The LIDAR distance measurements are published on two PointCloud topics on 30Hz each and IMU data is published by PX4 flight controller stack on 50Hz. The ground-truth pose data is extracted from the original dataset into a csv file and will be used for trajectory comparison in all scenarios. The original rosbag is then filtered to simulate a number of VL53L1X sensors with selected parameters. The filtered bags only contain the converted PointCloud2 and IMU besides the necessary clock messages.

The evaluation of Cartographer SLAM happens in incremental steps. It's unclear what parameters affect the performance the most, so instead of applying all of them at once, I chose to introduce these LIDAR parameters incrementally one by one to see the effects of them independently and work towards a complete set of parameters.

In the first simulation only converted but unfiltered range data is used, that has the maximum possible sampling rate, 4m maximum distance and 128x128 resolution in total of the top and bottom hemispheres. This will produce the most accurate map and localization estimates, it is unlikely that any down-sampled dataset will produce better results.

As the second step the effect of the resolution is tested. 13 LIDAR sensors are placed evenly on the horizontal axis and their resolution is set to 4x4 first and lowered to 3x3, 2x2 and 1x1. The sampling rate and maximum distance parameters are left unchanged.

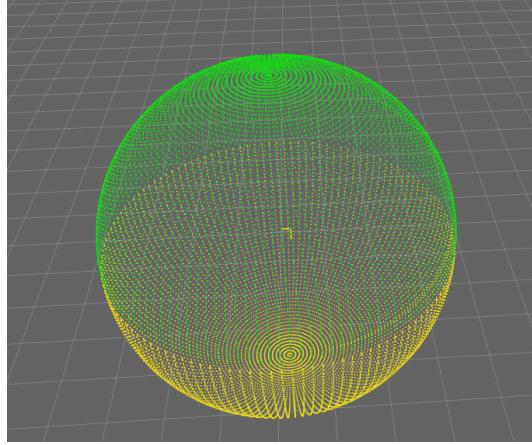
To investigate the effect of the sampling rate, the resolution with the best performance from the previous step is used, but with the sampling rate of the actual sensor. Finally the maximum distance is lowered, while keeping the same layout of the sensors.

After introducing all parameters the number of sensors is lowered until the lowest number sensors is found, that is still capable of producing a decent quality map.

### 4.2.1 Effects of LIDAR resolution

#### 4.2.1.1 Unfiltered data

In this step the range measurements inside the collected rosbag are converted to PointCloud2 without filtering and get written to a new bag. The LIDAR ranges visualized in Rviz can be seen on figure 4.11 to demonstrate the scan resolution.



**Figure 4.11:** Visualization of unfiltered LIDAR measurements maximized on 4 meters

Cartographer in 2D mode projects all points inside a scan to the horizontal plane before insertion and therefore it is not advisable to include points that are coming from ranges that are hitting the floor or the ceiling. Only points that are coming from reflections of objects should be included in the scans and saturated LIDAR measurements should be discarded. During the bandwidth filter step, Cartographer filters the points and only inserts the ones inside user set thresholds, others are discarded.

To exclude points coming from unwanted elevations or saturation, 4 parameters are adjusted in the configuration file: min\_z, max\_z to set the interval on z axis and min\_range, max\_range to set the range length interval. I have set min\_z and max\_z values to -10cm and 50cm, min\_range and max\_range to 40cm and 370cm accordingly. These settings proved to be working as expected and inserted ranges of the walls of the building. Since scans of the whole sphere are published on two topics, these need to be assembled into a single point cloud. This is done by Cartographer and is enabled by setting num\_accumulated\_range\_data to 2.

IMU usage is enabled using the use\_imu\_data parameter and IMU data will be used as an initial guess for the pose of the scan for insertion. Cartographer samples the accelerometer values for a certain period of time to determine the direction of gravity. The quadcopter changes pitch and roll angles quickly, so imu\_gravity\_time\_constant is set to 0.5 seconds to enable the algorithm to follow quick movements.

After setting up the basic settings for interfacing Cartographer, the actual tuning of the SLAM algorithm can be started. First the submap size need to be determined, that depends on the sampling rate of the LIDAR sensor. It is important because a new submap is built using constraints against the last two inserted submaps. Setting submap size low will cause submaps to drift or inadequate rotation due to lack of reference points. On the other hand setting submap size high doesn't allow global SLAM thread to rearrange submaps and calculate loop closures effectively.

For tuning local SLAM, global SLAM is disabled. In this case I found that lowering translation\_weight and rotation\_weight under ceres scan matcher options increase the mapping quality. The submap size is set to 90 ranges that is a bit too dense, but due to the high sampling frequency and resolution it still created an adequate map as seen on figure 4.12. Empty spaces are with white color and inserted scans are with black. The green points on the visualization are the points from the last inserted scan and the small coordinate systems are at the pose where a submap has been completed.

Corners are sharp, walls are straight and even small details like columns are visible on the final map. The big room on the top right of the map has three big windows and the drone was flying about in the height of them and that's the reason for the high quantity of empty space around this area. The algorithm was unable to insert points of the walls in between windows.

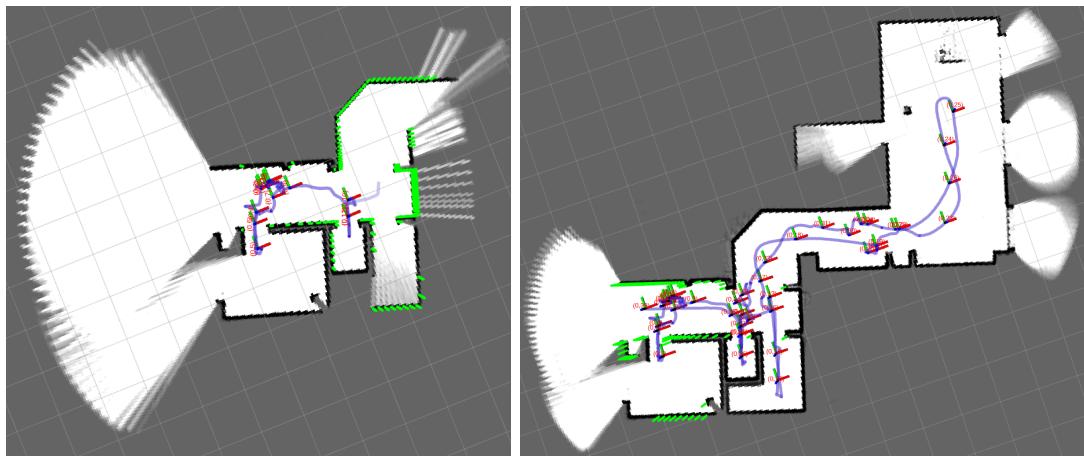
Because there is no noise added to ranges, even without global SLAM, a decent quality map can be built. If global SLAM is enabled, the linear\_search\_window parameter needs to be set to small, because the amount of drift of submaps is low. I have set the search window to 25cm, so submaps cannot be moved further than 25cm. If it is left unchanged, submaps in range of 15 meters can be pulled together and would mess up the map.

The extracted trajectory closely follows the ground truth trajectory as seen on figure 4.13. The RMS error on x axis is 6.3cm and a bit less 6.5cm on y axis, resulting in a cumulated 12.8cm RMS error.

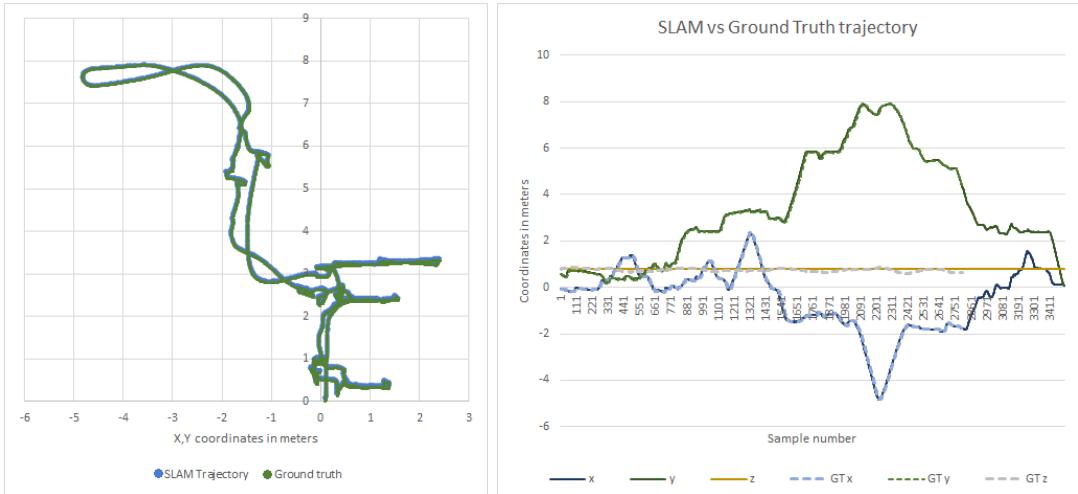
Figure 4.14 shows the absolute error on the axes separately. The error never goes above 25 centimeters on either axes and the fact that it always returns to 0 shows, that there is no drift in the pose estimation. It can also be seen on figure 4.13. Estimated and ground-truth coordinates overlap and produce a nicely overlapping trajectory as on the left image.

#### 4.2.1.2 2D SLAM with 4x4 resolution

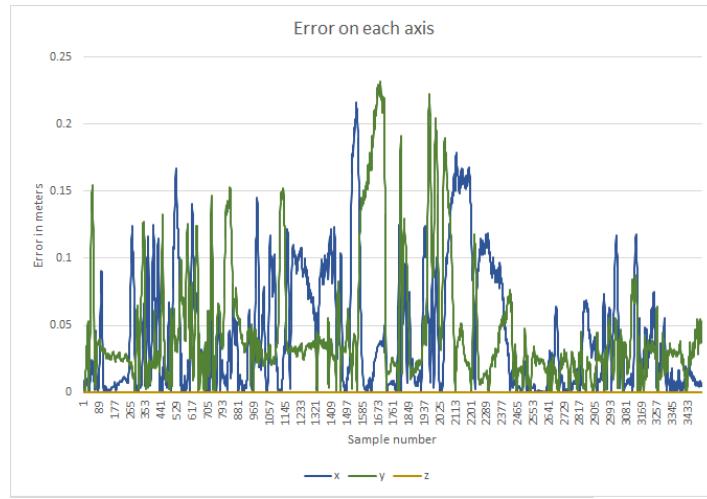
In this scenario, the original rosbag is filtered to simulate 13 LIDAR sensors placed evenly on the horizontal axis as described in 3.2, each with 4x4 resolution. The visualization of the point cloud from a single scan can be seen on figure 4.15.



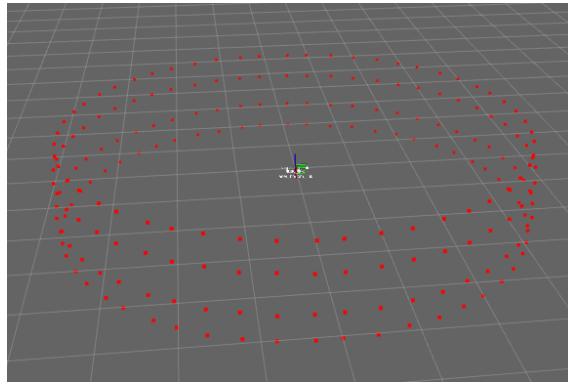
**Figure 4.12:** 2D SLAM map building on the left and complete map on the right



**Figure 4.13:** 2D SLAM extracted trajectory against ground truth



**Figure 4.14:** Trajectory error on each axis

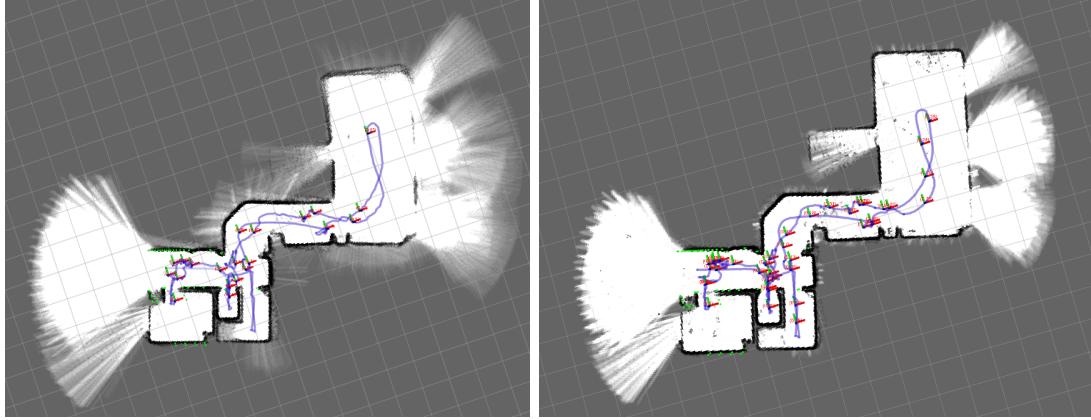


**Figure 4.15:** Visualization of ranges from 13 sensors with 4x4 resolution

The tuning of the SLAM algorithm with this setup wasn't more difficult, than it was with the previous setup, due to the high sampling rate and high number of points in each scan. For the local SLAM I have set the translation\_weight to 8 and rotation\_weight to 15. The reason for the increased weight is that otherwise the scan matcher algorithm was easily

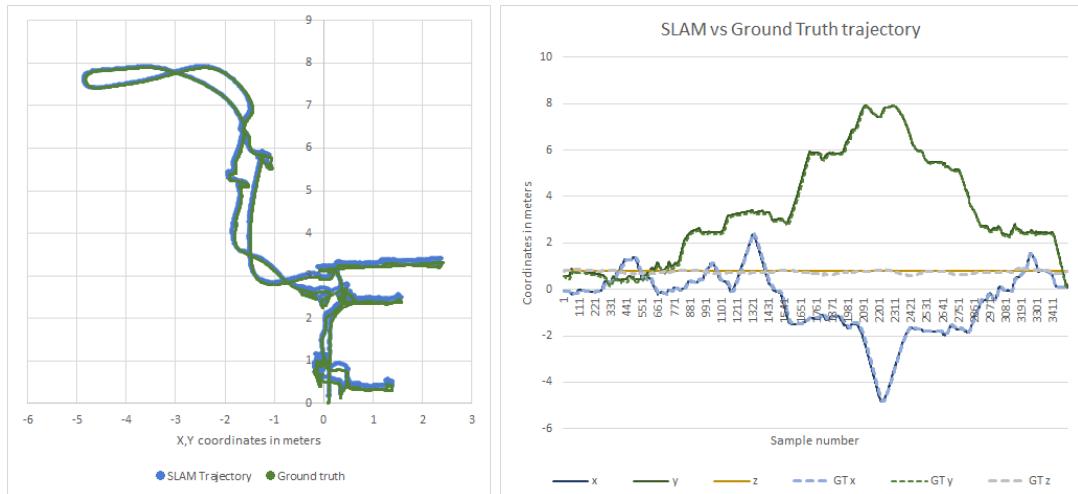
moving and rotating scans to unreasonable positions. With the higher weights the initial guess of the pose has higher influence.

To simulate the behavior of VL53L1X, ranges inside the sensor's field of view are averaged. This causes artifacts like curved corners and transition points that occur when an object is only partly in the field of view and another distant object covers the rest of the view. These smooth transitions hide the sharp edges of walls, making it harder to insert points and to find loop-closures.



**Figure 4.16:** Map before and after tuning using 4x4 resolution

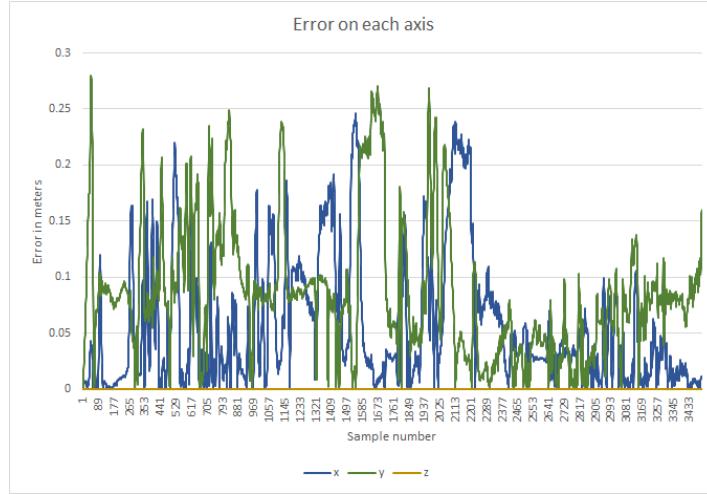
The final map after tuning is as seen on figure 4.16. The left image shows the built map before tuning and a failed loop-closure is seen in the top right of the map. The map on the right at first glance is accurate, there are no major drifts in either rotation or placement. Due to the averaging of field of view, corners are curved and the walls are not as sharp as they were using the unfiltered data. Transition points are best visible in the small room that opens from the big room. These are caused by the drone flying by the door and a part of a sensor's field of view hits the door post and the rest hits the wall behind. These ranges are averaged to a single point in between the two objects. These points are always present when the drone approaches a corner.



**Figure 4.17:** 2D SLAM extracted trajectory against ground truth using 4x4 resolution sensors

The estimated trajectory follows accurately the ground-truth trajectory as seen on 4.17. As seen on the left picture, there is an offset between the two paths and seems like a constant offset, but as seen on figure 4.18, there is no constant offset on neither axes.

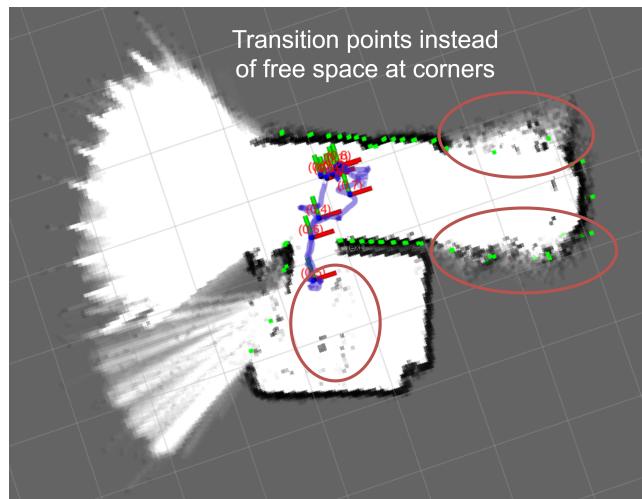
The RMS error of this setup is only 6cm more than the error of unfiltered data. The RMS error is 8.12cm on the x axis and 10.48cm on y axis, resulting in a cumulated error of 18.6cm.



**Figure 4.18:** Trajectory error on each axis using 4x4 resolution sensors

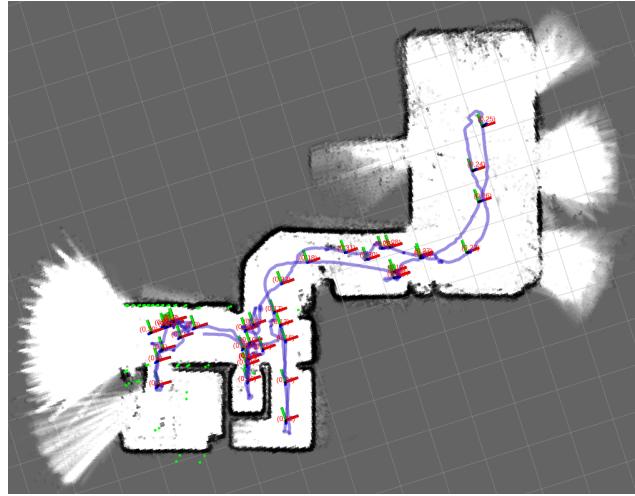
#### 4.2.1.3 2D SLAM with 3x3 resolution

As the resolution is lowered to 3x3, the field of view of each region of interest has broadened. This causes the artifacts coming from averaging to be more significant. An example for transition points can be seen on figure 4.19. The number of ranges per scan is also lowered, so the ratio of transition points and wall hits increases and it heavily impacts the performance of Cartographer.



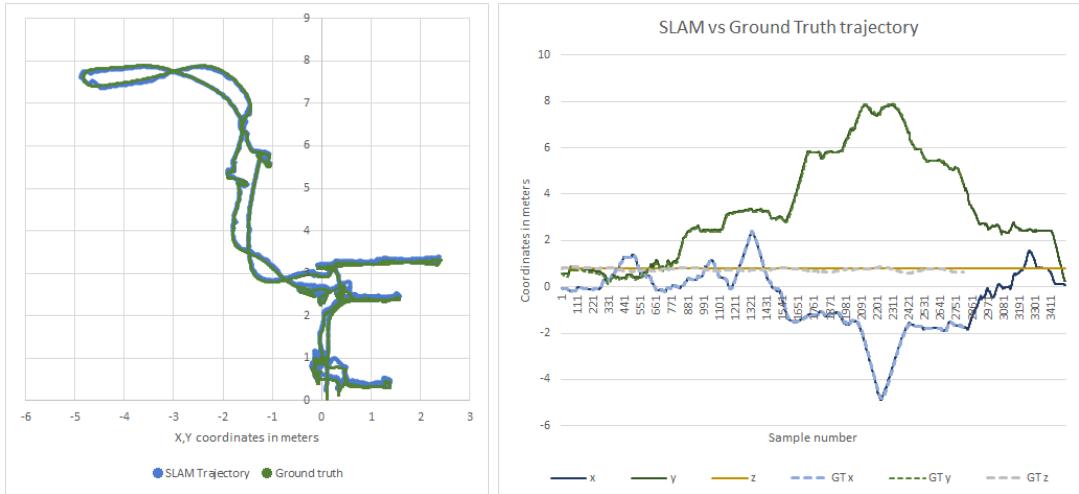
**Figure 4.19:** With wider field of view, the number of transition points increase

The tuned SLAM algorithm with sensors at a resolution of 3x3 seems to be stable and produces a decent quality map. On the way back submaps started to drift in orientation compared to the already inserted submaps, but global SLAM was able to compensate it and pulled submaps together. Another artifact is number of scans that could not be inserted properly into any submap and therefore produced a tilted shadow of the wall in the top right corner of the map.



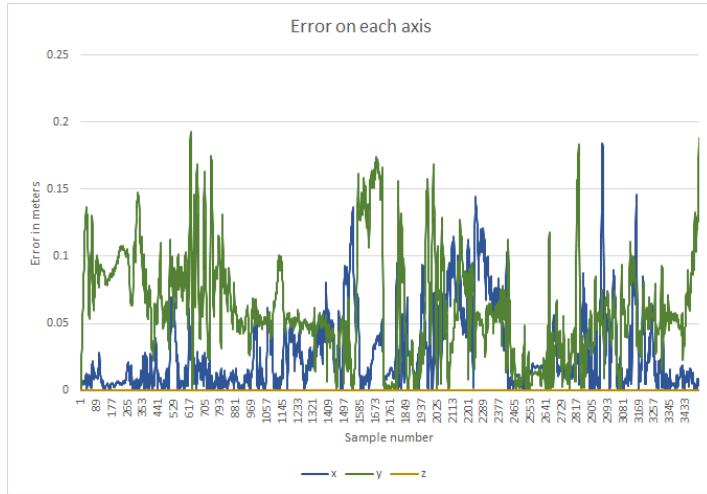
**Figure 4.20:** Map produced by sensors on 3x3 resolution

Although the mapping accuracy is reduced with the reduced sensor resolution, the localization accuracy has increased. The cumulated RMS error for this setup is 11.34cm. The RMS error on x axis is 4.04cm and on y axis it is 7.3cm.



**Figure 4.21:** 2D SLAM extracted trajectory against ground truth using 3x3 resolution sensors

The accuracy of this setup is unexpected, because it contains less points per scan and more the presence of transition points is significantly higher. I think the higher accuracy is due to the better tuning of Cartographer, because and shows the flaw of manual tuning and visual inspection. Also shows the importance of tuning.

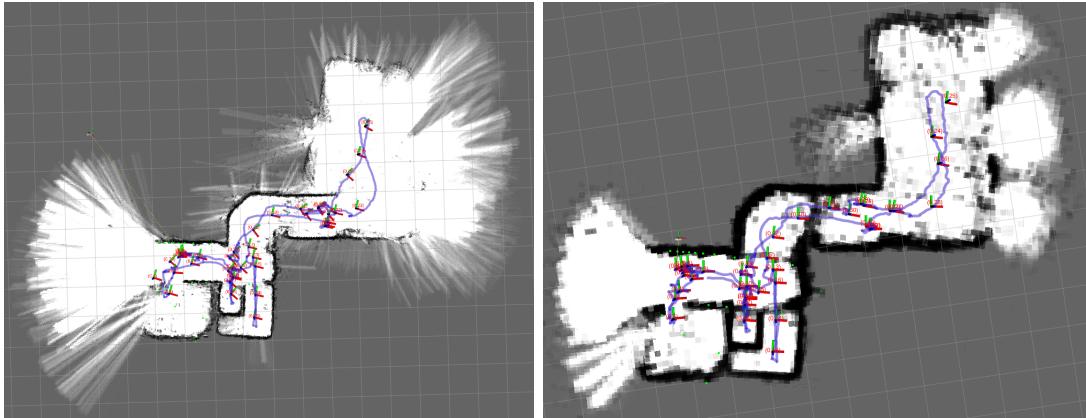


**Figure 4.22:** Trajectory error on each axis using 3x3 resolution sensors

#### 4.2.1.4 2D SLAM with 2x2 resolution

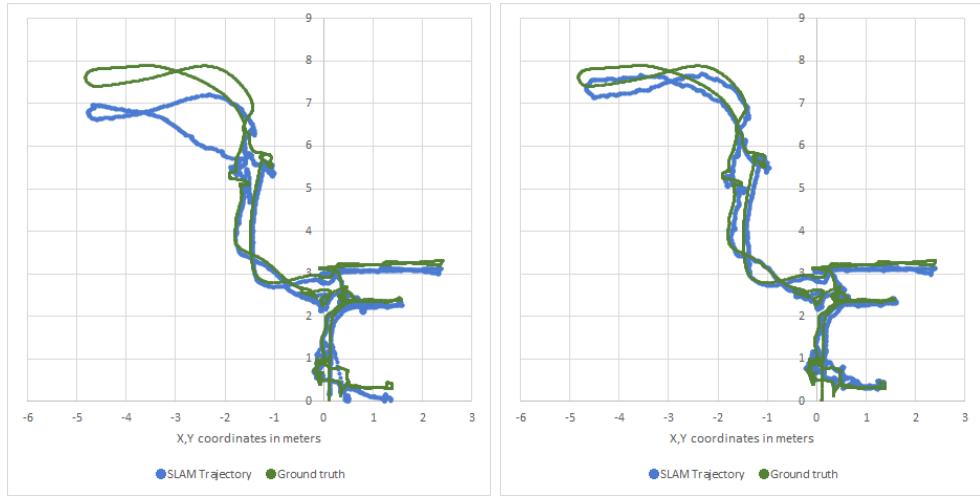
Compared to 3x3 resolution, in 2x2 each scan contains 66% less points. A complete scan of 13 sensors consist of only 52 points that makes scan matching significantly less efficient. Tuning of this setup was noticeably harder and the quality of the produced map has highly decreased.

To compensate the scarce ranges inside scans, I have also run the simulation with lowered grid resolution from 5cm to 15cm. The difference of the two produced maps is seen on figure 4.23. The lowered resolution resulted in a more consistent map, but the transient points are also better visible on this map.



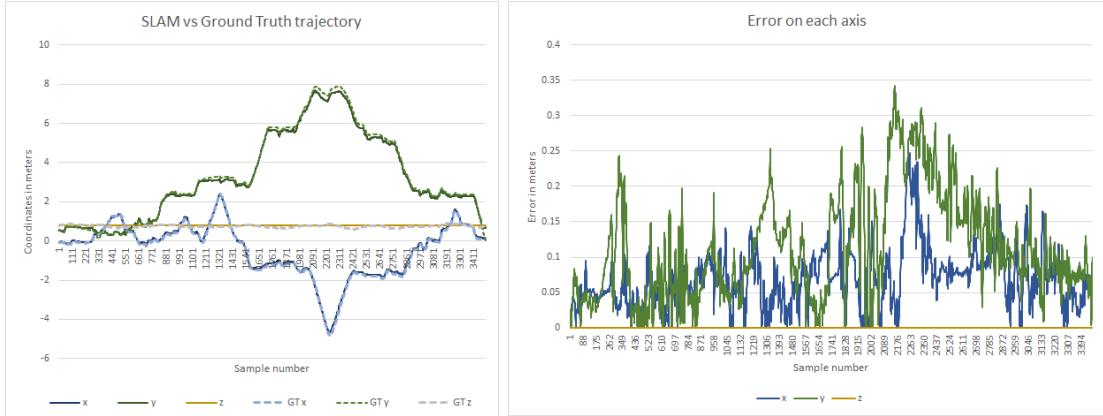
**Figure 4.23:** Produced map using 5cm grid resolution on the left and 15cm on the right using 3x3 resolution sensors

By placing the trajectories side by side as on figure 4.24, the difference is can be easily seen. While the configuration with high resolution grid map is able to follow the ground-truth trajectory in small rooms and the narrow corridor, it loses track in the big and spacious room, when the ranges are far apart. Lower resolution makes the SLAM algorithm more efficient mostly in the big room, but also has less drift and it is more stable.



**Figure 4.24:** Comparison of estimated trajectories using 5cm grid resolution on the left and 15cm on the right

The high resolution trajectory is at some points as far as 1.4m from the ground-truth pose, but the cumulated RMS error is only 43.4cm. This shows that a seemingly small difference in RMS error can mean an error of more than a meter at a part of the trajectory. The low resolution trajectory in comparison has a cumulated RMS error of 22.13cm. 7.95cm error on x axis and 14.17cm on the y axis.



**Figure 4.25:** 2D SLAM extracted trajectory against ground truth using 2x2 resolution sensors and 15cm grid resolution

#### 4.2.1.5 2D SLAM with 1x1 resolution

By further lowering the resolution, each sensor has a single range measurement as output resulting in 13 range measurements per a complete scan. Such a low resolution hides most of the details and almost always has multiple walls in its field of view. For example on 4m distance with a 27° angle, the field of view opens to 1.8m. All ranges inside this wide field of view get averaged resulting in a single distance measurement. On the corridor the drone flies in, ranges of a single sensor's scan can hit the two parallel walls and also the wall at the end of the corridor. This scan will produce a floating point in the middle of the corridor.

Due to the high rate of the above mentioned artifacts, I was not able to tune the SLAM algorithm. With all settings tried, the system proved to be instable and unreliable.

#### 4.2.1.6 Summary

After running and tuning the SLAM algorithm on all resolutions, I can conclude that higher resolutions positively affect the stability and reliability of Cartographer. Lower resolution means broader field of view that hides details like corners and edges, resulting poorer quality maps.

Table 4.1 shows the the RMS error for each resolutions I have tried. Surprisingly the setup with 3x3 sensors has a smaller RMS error, than the unfiltered setup or 4x4 setup. The reason for this might be because I have conducted these experiments in the same order as in this document and I have gained more experience in tuning of Cartographer by the time I have got to the experiment with 3x3 resolution sensors.

	Unfiltered	4x4	3x3	2x2	1x1
RMS error on x axis	6.3cm	8.12cm	4.04cm	7.95cm	-
RMS error on y axis	6.5cm	10.48cm	11.34cm	14.17cm	-
Cumulated RMS error	12.8cm	18.6cm	15.74cm	22.13cm	-

**Table 4.1:** RMS error on different resolutions

#### 4.2.2 Effects of sampling rate

In the investigation of the effects of sampling rate, only two resolutions with the lowest RMS error are used. These resolutions are 4x4 and 3x3. In this test, the same LIDAR layout is kept as before, 13 sensors evenly placed on the horizontal axis. Only one parameter is changed, that is the sampling rate of each sensor.

For 4x4 resolution I chose the sampling time at 400ms, based on the measurements in section 4.1.2. I chose the setup with a timing budget of 33ms over 18.5ms, because it has a lower sampling rate and if it works using this setting, it should be working with higher sampling rate too.

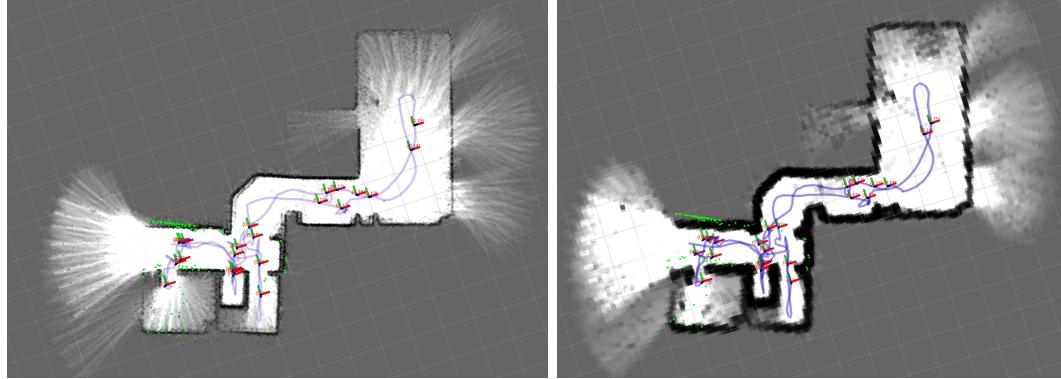
As for 3x3 resolution I chose the setup with the same sampling rate of 33ms. This setup has a sampling time of 220ms.

##### 4.2.2.1 4x4 resolution

The dropped sampling rate from 30Hz to 2.5Hz is significant and makes tuning much harder. In previous simulations, the occupied\_space\_weight parameter was not at all or only slightly changed, but in this case this parameter proved to be the key parameter. This parameter sets how heavily Cartographer should rely on already inserted points. The higher this value, the less flexible already inserted points are and are less likely to be overwritten by new scans.

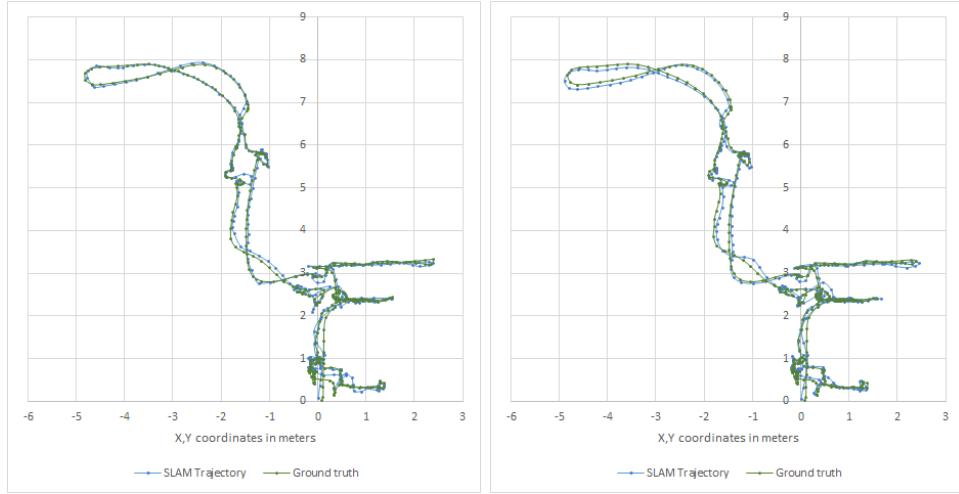
As seen in the previous simulation that uses 2x2 resolution sensors, I have also tried in this case to lower the grid resolution to 15cm. The two completed maps are seen on figure 4.26. The quality of the map is decent in both cases, no major flaw is visible. In both cases global SLAM had to compensate rotation drifts, that occurred when the drone turned

around in the big room. This was in general a crucial point in the tuning procedure, most often submaps fell apart at this point.



**Figure 4.26:** Maps built using 4x4 resolution sensors. 5cm grid resolution on the left and 15cm resolution on the right

The two trajectories 4.27 look similar, both follow the ground-truth trajectory closely. Low grid resolution produces a smoother path, while on high resolution there are some edges and sharper turns. In general the configuration with 5cm grid resolution felt very unstable and that it can fall apart at any point. If submaps are built properly but misplaced relative to each other, it can still be compensated by global SLAM, but in this case submaps fell apart, that cannot be compensated anymore.



**Figure 4.27:** Extracted trajectories using 4x4 resolution sensors. 5cm grid resolution on the left and 15cm grid resolution on the right

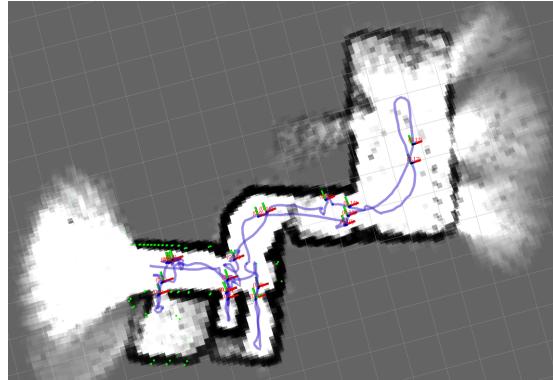
The cumulated RMS error on 5cm grid resolution is 16.49cm and on 15cm resolution it is higher: 18.95cm. Higher resolution grid results in greater accuracy in position estimates, but its stability is lower in comparison to lower resolution.

#### 4.2.2.2 3x3 resolution

The benefit of using 3x3 resolution setting for the LIDAR is the higher sampling rate. Using the same timing budget of 33ms as for 4x4 resolution, the sampling time is 220ms

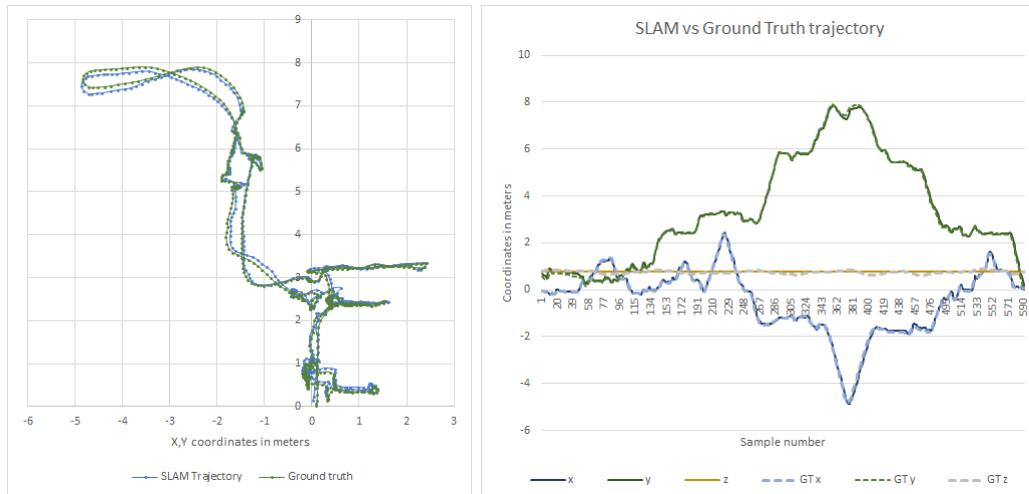
in comparison to the previous 400ms. This results an 80% increase in update rate to 4.5Hz, that also means almost twice more often pose updates.

Due to high instability I was unable to tune SLAM with 5cm grid resolution. Lowering the grid resolution once again to 15cm, resolved the issue and the system felt reliable afterwards. The produced map using higher grid size is on figure 4.29. The map quality is adequate, global SLAM almost haven't done any corrections.



**Figure 4.28:** Map built using 3x3 resolution sensors and 220ms sampling time

The extracted trajectory is seen on figure 4.29. The cumulated RMS error is 14.67cm, 6.4cm on x axis and 8.2 on y axis. Based on the stability and also accuracy 3x3 resolution seems like a better resolution for Cartographer SLAM.

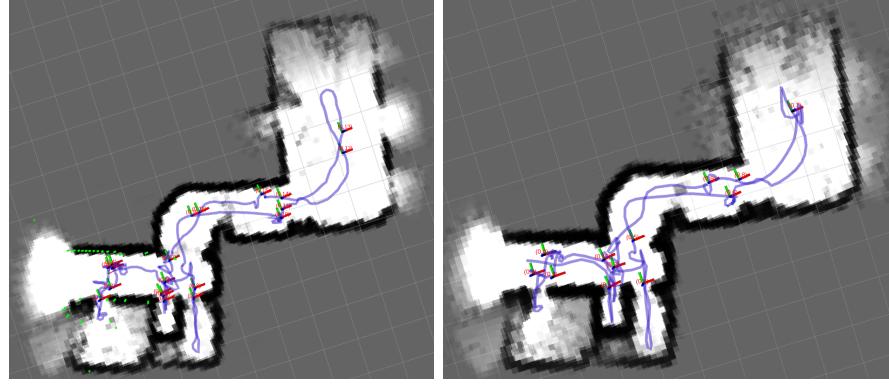


**Figure 4.29:** Extracted trajectory using 3x3 resolution sensors and 220ms sampling rate

### 4.2.3 Effects of maximum distance

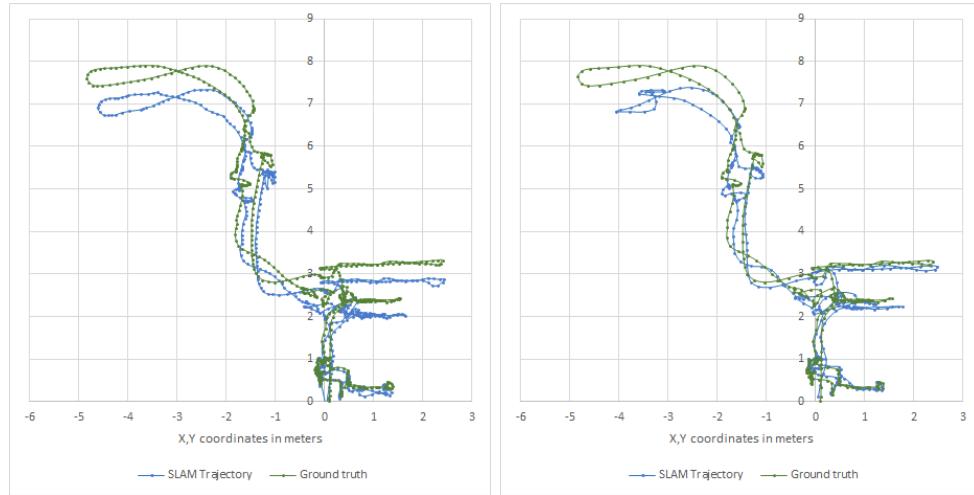
The measurements with VL53L1X sensor showed that the each setup works accurately until 2.5m and each has about 25cm drift at distance of 3m. I chose 2.75m as maximum distance for both 3x3 and 4x4 resolution sensor simulations, because this is the distance reported by the sensor on 3m.

The lowered maximum distance had a positive effect, that the number of transition points has been reduced. This is because the shorter distance means that the field of view is shortened and cannot open as wide as it could before. The downside of the limited range is that the algorithm loses track in the big room, because it's longer than 5m and the two farther sides do not fit in a single scan. This caused more flaws on 4x4 resolution, while the other map looks similar to the one before the introduction of the distance limit. A small tilt can be observed that there is a slight tilt between the room on the right and the rest of the map. The produced maps using both 3x3 and 4x4 resolution setups are on figure 4.30.



**Figure 4.30:** Maps produced by SLAM algorithm with 2.75m maximum distance and realistic sampling rate enabled, 3x3 resolution on the left, 4x4 resolution on the right

The trajectory using sensors with 3x3 resolution has a significant offset on y axis, that appeared after the drone exited the very first room. This offset is picked up, because with maximized distance the corridor appears to be open on both ends and without any reference points. Similar scans are pulled together, resulting in a shortened corridor.



**Figure 4.31:** Extracted trajectories with 2.75m maximum distance and realistic sampling rate, 3x3 resolution on the left and 4x4 resolution on the right

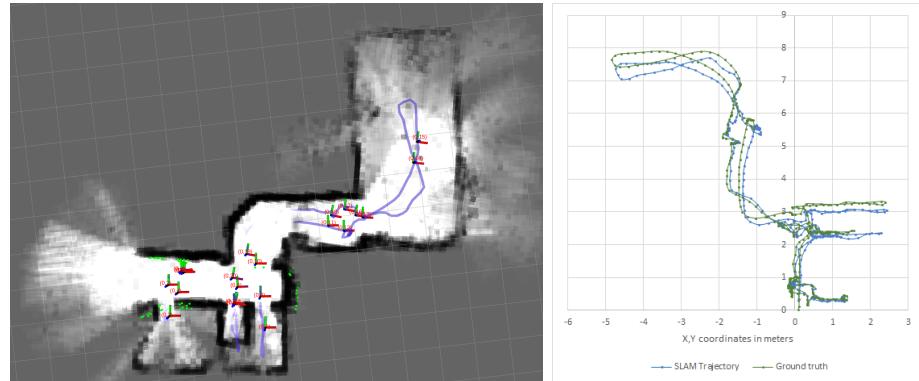
The trajectory using the setup with 4x4 resolution sensors has a smaller offset. Seems like with more points per scan the SLAM algorithm was able to pick up small details

and better track the movement of the drone on this mostly even corridor. On the other hand it failed in the big room, where the same situation appears. The stairs at the end of the room also make tracking harder, ranges from the stairs and the wall are hard to differentiate.

The cumulated RMS error for 3x3 resolution is 45cm and 49.6cm for 4x4. In the end the mostly constant offset with otherwise correct shape results in a smaller error, than the pose estimation that follows ground-truth closely, but fails in shape at a relatively small area.

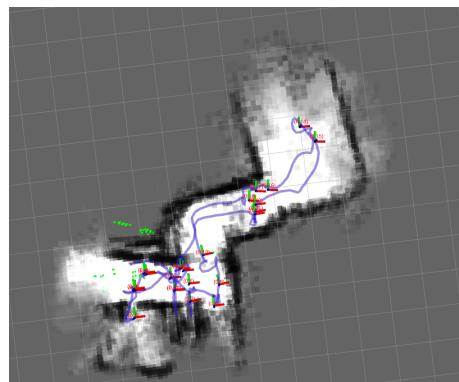
#### 4.2.4 Reduced set of evenly distributed LIDAR sensors

The goal is to find the minimum number of LIDAR sensors, with which the SLAM algorithm still works reliably and can produce a coherent map. To see the effect of reduction in the number of sensors, I have decided to halve the set and test the performance using only 6 sensors. The sensors are evenly placed to cover the 360° horizontal plane.



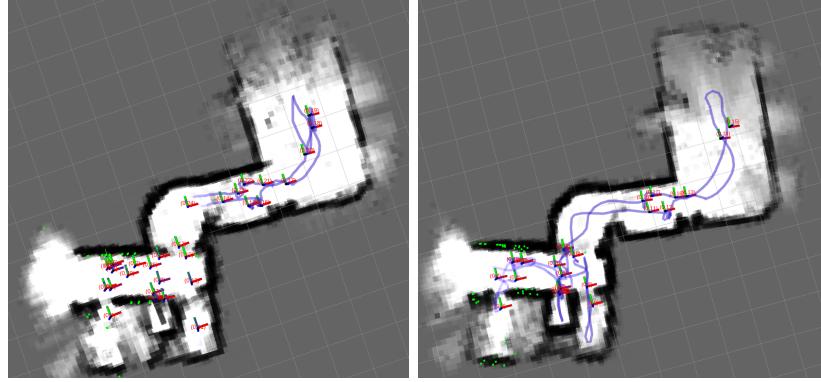
**Figure 4.32:** Map and trajectory using 6 pieces of 4x4 resolution sensor with 400ms sampling time and 4m distance limit

At first I used 4x4 resolution sensors with 400ms update rate and without limiting the maximum distance. This means the maximum distance of each sensor is 4m. The produced map is almost as good as it was with 13 sensors, but tuning was difficult and it was sensitive to any small changes in weights. The cumulated RMS error of this setup is 37.32cm.



**Figure 4.33:** Failed mapping attempt with 4x4 resolution, 400ms sampling time and 2.75m maximum distance

After tuning the setup without distance filter, I have used the same number of sensors and layout with distance filter enabled. The maximum distance is set to 2.75m as in previous simulations. The tuning of the algorithm proved to be significantly harder this time. In the end I couldn't find any combination of parameters that would enable the mapping to work reliably. Due to the low number of sensors, there are not enough remarkable details in scans that would enable proper scan matching. The best produced map using 4x4 resolution sensors is on figure 4.33. Similar maps were created using 3x3 resolution.

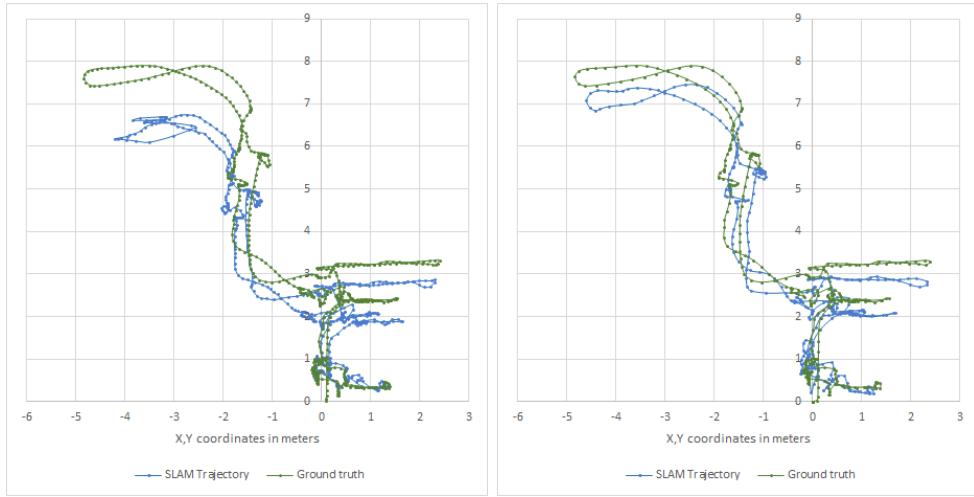


**Figure 4.34:** SLAM maps using 8 sensors in even layout, realistic sampling time, 2.75m distance limit, 3x3 resolution on the left and 4x4 on the right

After the failed attempt using 6 sensors, I have raised the number to 8 and filtered the original bag using 3x3 and 4x4 resolution. The increased number had a positive effect and I was able to tune the algorithm properly. In contrary to the measurements in 4.2.3, this time setup using 4x4 resolution successfully tracked the drone even in the big room, while 3x3 resolution failed.

RMS cumulated error using 8 sensors with 4x4 resolution is 34.1cm and almost triple 89.5cm using 3x3 resolution. The extracted trajectories can be inspected on figure 4.35. In different quantities, but both trajectories have an offset that cumulated after exiting the first room, due to the shortening of corridors. This offset persists throughout the trajectories afterwards, but the submaps are properly aligned relative to each other.

I have tried to lower the number of sensors to 7, but I have got the same results as with 6 sensors. Therefore the setup using 8 sensors is assumed to be the lowest number of VL53L1X sensors, that is still able to produce consistent maps with Cartographer SLAM in 2D mode.



**Figure 4.35:** SLAM trajectories using 8 sensors in even layout, realistic sampling time, 2.75m distance limit, 3x3 resolution on the left and 4x4 on the right

### 4.3 3D SLAM evaluation

To evaluate the 3D SLAM performance, I decided to work towards the minimum number of LIDAR sensors in incremental steps, just like in case of 2D. 3D SLAM is an extension of 2D SLAM and most concepts learned in 2D apply to it. Nevertheless 3D SLAM is significantly more complex and tuning procedure is harder due to the larger number of configuration parameters. Rviz can only visualize the 2D projection of the 3D map, this will be used as a base for tuning.

For the first simulation, I have converted the PointCloud messages, but left data unfiltered to see the maximum accuracy of the 3D SLAM algorithm using unfiltered data.

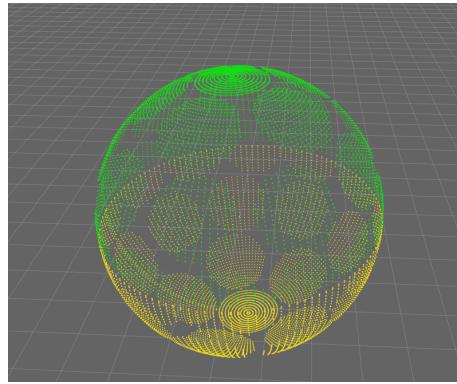
To see the effects of resolution, sampling rate and maximum distance I have used 43 sensors with 4x4 resolution and introducing the according parameters incrementally. In the layout 13 sensors pointing at  $0^\circ$ , 9-9 at  $\pm 30^\circ$ , 5-5 at  $\pm 60^\circ$  and 1-1 at  $\pm 90^\circ$  all relative to the horizontal plane and evenly placed around the vertical axis as on figure 4.36.

The 3D SLAM evaluation happens in incremental steps. All steps will be tested using 4x4 sensor resolution setting, because this resolution is expected to have the best performance. After completing all the steps, 3x3 resolution is also evaluated with all parameters introduced at once. Based on the results of the previous simulations, I try to lower the number of sensors until the produced map is still consistent.

#### 4.3.1 Unfiltered data

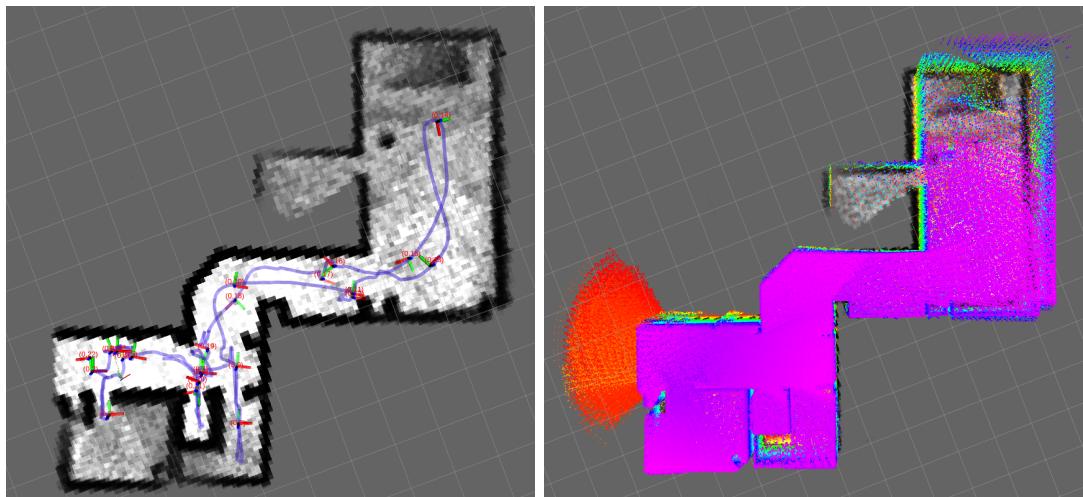
As the first step, the same settings need to be set up as used in 2D SLAM. These settings are the minimum and maximum range, number of range data to accumulate, IMU gravity time constant, but there is no option to exclude ranges based on elevation.

Cartographer in 3D SLAM mode uses two probability grid resolutions to reduce computation complexity, a low and a high resolution. Cartographer in general is optimized for LIDAR sensors with a range up to and above 100m, therefore the low resolution grid has a large grid size of 45cm. To make it work more accurately with the selected VL53L1X sensor, I have lowered the low resolution grid size to 15cm instead. Finding the right ro-



**Figure 4.36:** Visualization of ranges inside field of views of 43 sensors without averaging applied

tation is a computationally expensive operation for the scan matching algorithm, and for this reason the rotation\_weight parameter was set to 400. The drone has quick rotations, so the algorithm needs to compensate these rotations. To allow rotations I have lowered the rotation\_weight parameter and found 15 to be an optimal value.

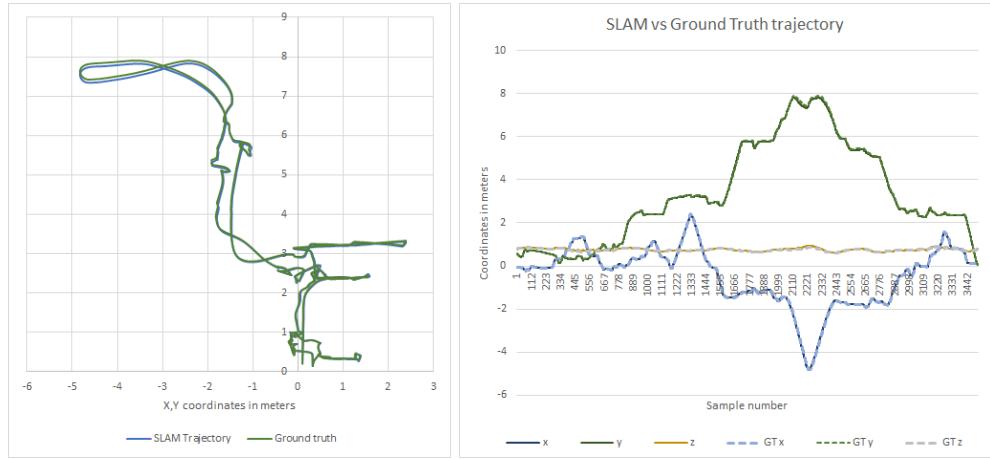


**Figure 4.37:** 3D slam map and inserted points using unfiltered data

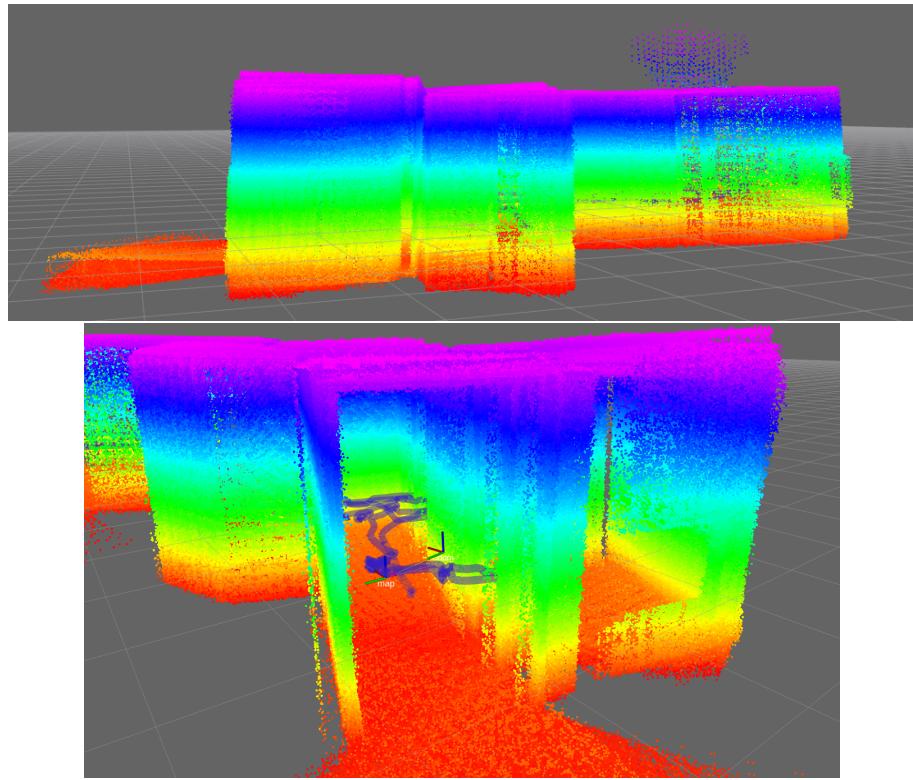
The estimated trajectory against ground-truth is visible on figure 4.38. The error of the estimation is relatively small and it typically stays under 15cm, but never goes above 30cm. The cumulated RMS error for the unfiltered setup is 10.4cm, with 3.6cm error on x, 4.4cm on y and 2.4cm on z axis.

The visualization of the point cloud on these images happened, by setting the decay time parameter to a high number in Rviz. This way all matched points published by Cartographer will be displayed in this visualization. These point clouds however do not change with optimization of global SLAM, but stay fixed as the submap was originally inserted. Two more images can be seen of the visualization on figure 4.39.

Some tilt can be observed relative to the horizontal plane. To see if this tilt is persistent or has been corrected, the final optimized point cloud can be extracted using Cartographer's asset writer functionality. After extraction the cloud is visualized using Point Cloud Viewer developed by Google.



**Figure 4.38:** 3D slam trajectory using unfiltered data



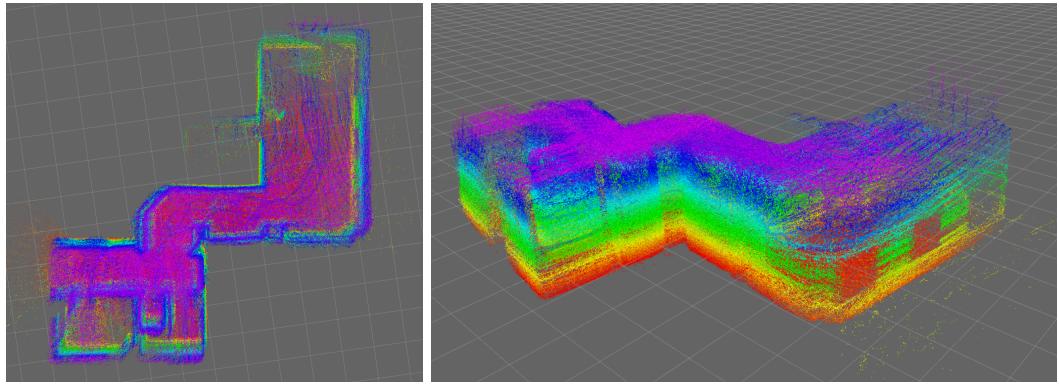
**Figure 4.39:** 3D slam visualization of inserted points in Rviz

TODO Point Cloud viewer image

#### 4.3.2 Effects of LIDAR resolution

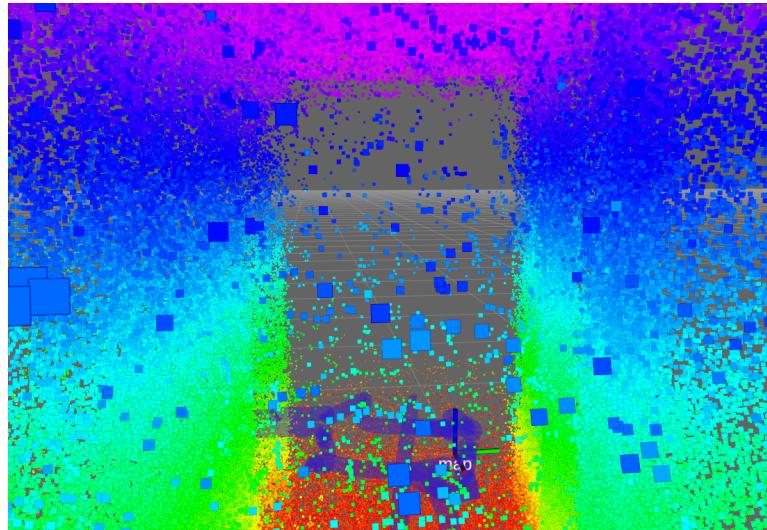
In this setup my goal is to see what are the effects of the lowered resolution on 3D SLAM. I have evenly placed 43 sensors with 4x4 resolution to cover the whole sphere. The update rate and maximum distance parameters are left unchanged at 30Hz and 4.0m respectively.

The resulting point cloud is seen on figure 4.40. Due to the averaging, the corners became curved, but all in all the algorithm was able to follow the drone accurately and map its surroundings. The stability was unchanged relative to the unfiltered data.



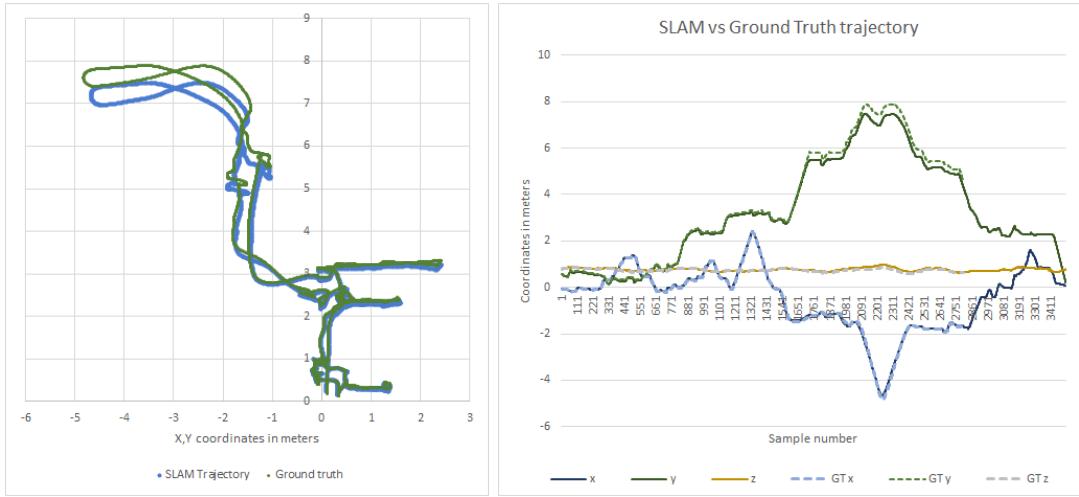
**Figure 4.40:** Point clouds in Rviz produced by 3D SLAM using 4x4 resolution sensors

The artifact of transition points is better visible on 3D visualization on figure 4.41. These floating points are caused by the wide field of view and makes it harder for the scan matching algorithm to find scan matches.



**Figure 4.41:** Transition points inserted due to wide field of view

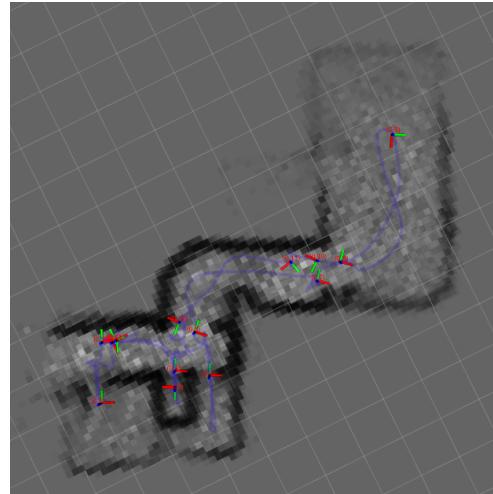
The estimated trajectory follows the ground-truth trajectory closely until the drone enters the corridor close to the big room. The trajectory in this corridor is shortened resulting in an offset in the following poses. The cumulated RMS error is 29.98cm for this setup.



**Figure 4.42:** 3D trajectory against ground truth using 4x4 resolution sensors without limiting maximum distance or sampling rate

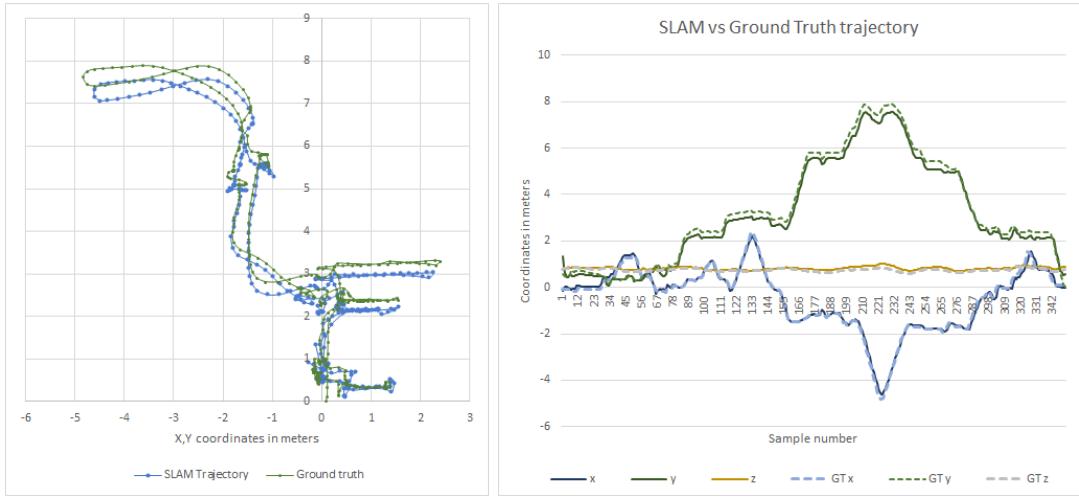
### 4.3.3 Effects of sampling rate

To see the effects of lowered sampling rate I have kept the layout from previous simulations and applied sampling rate filter with a sampling time of 400ms. The produced map is on figure 4.43. A tilt can be observed on the bottom left of the map, but otherwise the map is consistent and was able to track the drone even in the big room.



**Figure 4.43:** Map produced by 3D SLAM using 4x4 resolution sensors and 400ms sampling time

The trajectory using 4x4 resolution sensors and sampling rate filter can be seen on figure 4.44. As seen in previous measurements, the corridor is shortened when the drone exits the first room. In addition to this offset, the second corridor before the big room is also shortened that results in about 35cm of cumulated offset error that persists until the drone returns to the starting position. The cumulated RMS error using this setup is 43.87cm.

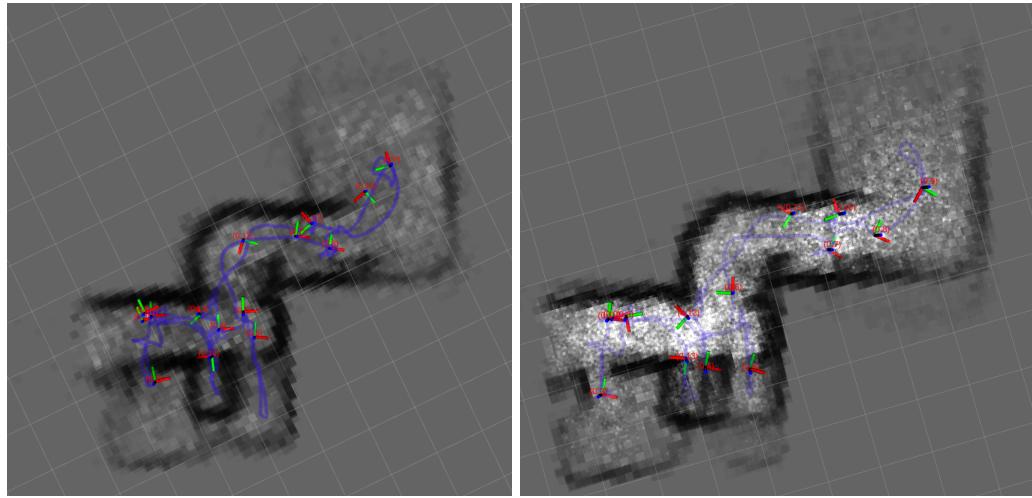


**Figure 4.44:** 3D trajectory against ground truth using 4x4 resolution sensors with sampling rate filter applied

#### 4.3.4 Effects of maximum distance

By enabling maximum distance filter, all parameters are enabled for the simulation of VL53L1X LIDAR sensors. The layout, resolution and sampling time is left unchanged compared to the previous measurement and the maximum distance is set to 2.75m.

The tuning of SLAM has proved to be significantly harder, than it was with any setups before. The algorithm was unable to track the position of the drone reliably and produced poor quality maps with all configurations I have tried. In the tuning process I have tried multiple combinations of grid sizes for low and high grid resolution, experimented with weights for rotation, translation and occupied space weight, lowered voxel filter size and tried different submap sizes.



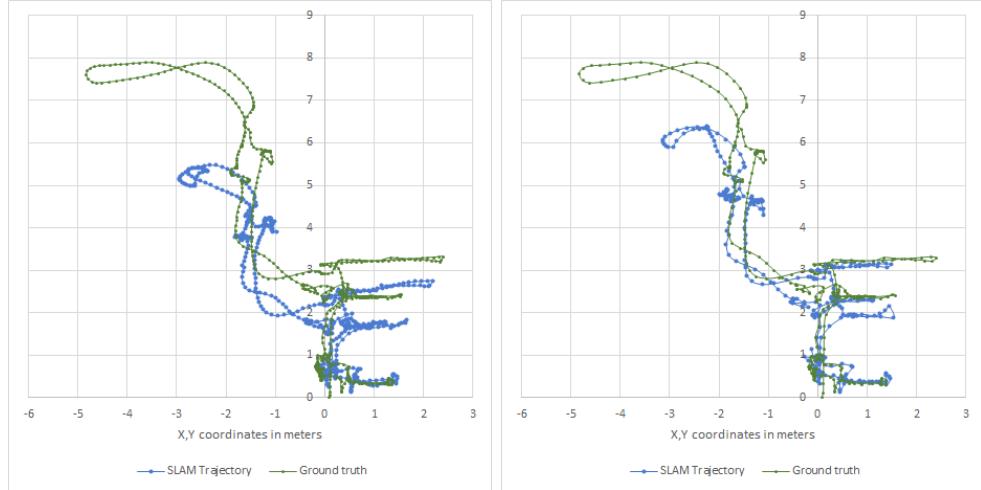
**Figure 4.45:** Map produced by 3D SLAM using all simulated parameters, using 3x3 resolution on the left and 4x4 resolution on the right

After extensive research on the tuning process, I still couldn't find a configuration that would allow the algorithm to produce decent quality submaps. To see how a higher sampling rate affects the mapping quality, I have lowered the resolution to 3x3 and set

the sampling time to 220ms. The mapping quality hasn't improved significantly. The best quality maps using these setups are seen on figure 4.45.

I have conducted multiple experiments using different sensor layouts. I have tried to shift the rows of sensors relative to other rows and also tried to remove the top and bottom sensors to see if that led to incorrect scan matching. I have found no relation between the sensor layout and the SLAM performance in these cases.

The extracted trajectories relative to the ground-truth data have 113.34cm RMS error in case of 4x4 resolution and 163.71cm for 3x3. The trajectories are on figure 4.46.



**Figure 4.46:** Trajectories produced by 3D SLAM using all simulated parameters, using 3x3 resolution on the left and 4x4 resolution on the right

#### 4.3.5 Summary

During the evaluation of 3D SLAM, I have incrementally included each parameter to the evenly layout LIDAR sensors. Lowering the resolution to 4x4 had a increase in RMS error, but the algorithm was able to track the drone and produce reliable maps. With the application of the sampling rate filter, the RMS error further grew, but was still able to produce adequate maps. On the other hand the introduction of maximum distance filter has completely changed the behavior of Cartographer SLAM and I was unable to properly tune it. Changing the sensors to 3x3 resolution to increase the sampling rate, had no positive effects on the SLAM performance.

	Unfiltered	4x4	4x4,400ms	4x4,400ms,2.75m	3x3,220ms,2.75m
X	3.6cm	5.6cm	11.2cm	39.75cm	37.25cm
Y	4.4cm	19.8	24.3cm	69.75cm	122cm
Z	2.4cm	4.4cm	8.37cm	3.84cm	4.4cm
Cumulated	10.4	29.98cm	43.87cm	113.34cm	163.7cm

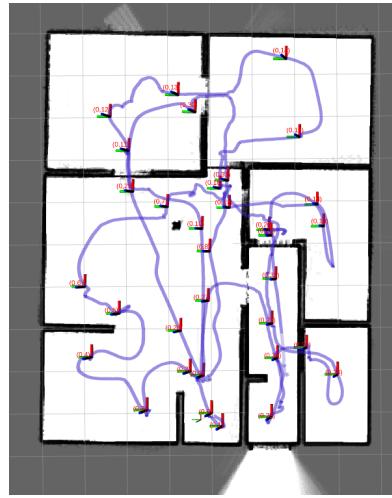
**Table 4.2:** RMS error from setups used for 3D SLAM evaluation

During the above described experiments, I could not make Cartographer work in 3D SLAM mode even using 43 sensors in total. These experiments were conducted in Gazebo simulator and no additional noise is included. In actual builds, the performance is expected to be

worse. To compensate, adding more sensors would most likely improve the performance of the SLAM algorithm, but placing 43 sensors on a real-world quadcopter is already not a feasible project.

## 4.4 SLAM performance on different building

So far the system implementation has only been tested on a single building and using the same dataset. To see the performance of the system, I have created a new building with different floor map as seen on figure 4.47. The building has rooms with varying size and shape to test SLAM performance in different scenarios.

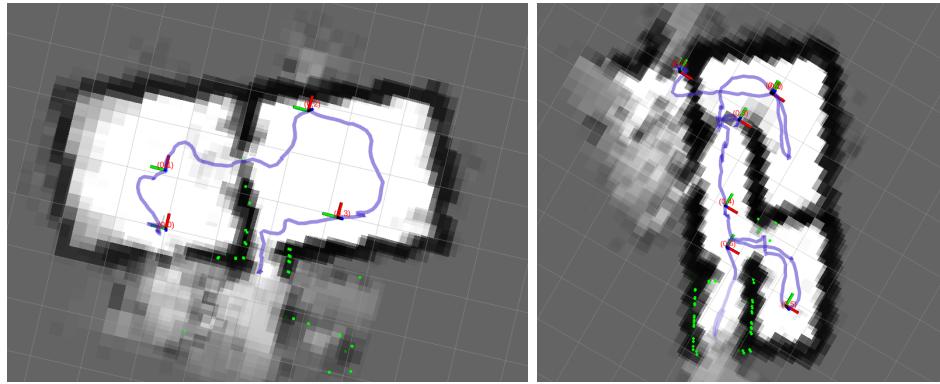


**Figure 4.47:** Floor map of second building built using unfiltered data

As described at the end of the previous section, 3D configuration doesn't work reliably, therefore only 2D configuration will be used in this section. As for sensor settings I chose the configuration with 3x3 resolution, because it provides higher update rate while it has similar performance to the 4x4 resolution alternative.

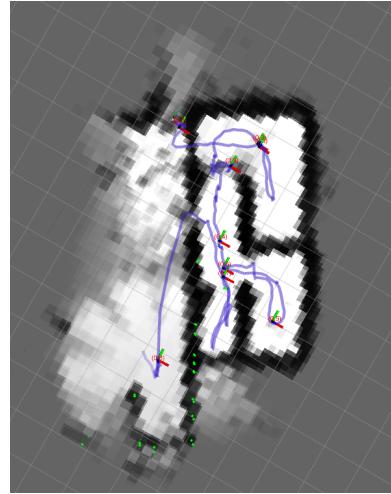
During the tuning I found that the SLAM setup works well in smaller rooms, but loses track in the big room, that messes up the previously built map. Even global SLAM is incapable of correcting these errors, because these are contained inside submaps and not between submaps.

To determine the limitations of the proposed LIDAR setup and Cartographer SLAM, I broke down the collected data to subsets that contain measurements only of the corridor and small rooms on the right, the top two rooms and the big room. Figure 4.48 shows the mapping performance using data of the small rooms and the corridor. The other image shows the map of the top medium sized rooms. Both maps are decent quality and the algorithm was able to track the drone during these flights.



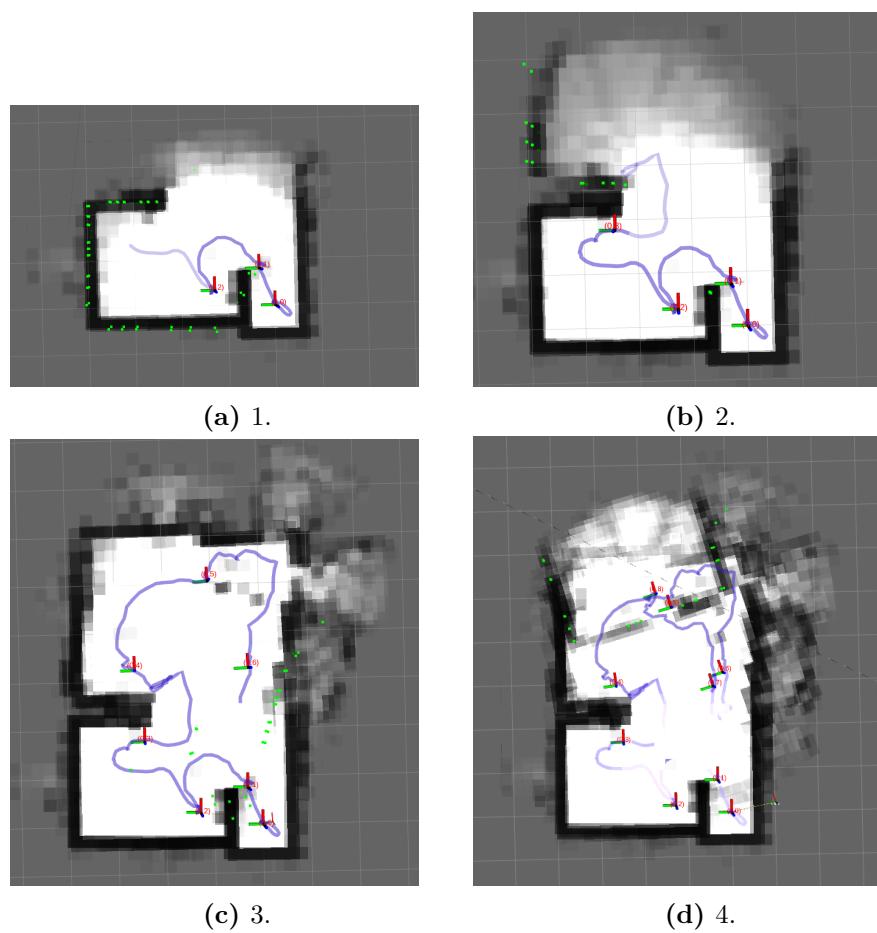
**Figure 4.48:** Mapping performance in smaller rooms

The next image on 4.49 shows the transition from the corridor to the main room. The maximum range of the sensors is 2.75m and the room is about 5m wide and almost 7m long, therefore the LIDARs fail to pick up signal from both sides of the room in the same time. Due to the lack of reference a strong slip appears in the produced map, about 3m.



**Figure 4.49:** Mapping performance in smaller rooms

Figure 4.50 shows the mapping process in 4 steps. The mapping quality is adequate in confined spaces like on subfigures 1 and 2, but some slippage can be observed when the drone flies around the wall in between two bay areas. The slippage becomes even greater when the drone flies out of the second bay into a wide open space as seen on the third subfigure. At this point a tilt can also be observed and the map becomes completely messy from this point. Submaps of the small rooms are still correctly built, but are misplaced due to the error in the main room.



**Figure 4.50:** SLAM mapping process in open space with additional details

# Chapter 5

## Conclusion

In this thesis I have researched the topic of indoor SLAM application using a set of cheap single-point LIDAR sensors. I have used data collected from a single simulated flight to evaluate different sensor layouts and settings, by filtering this flight data to simulate a certain number of VL53L1X sensors and their parameters.

I have measured the performance of VL53L1X sensor using different operating modes. I have determined the maximum ranging distance, update rate and noise level of each operating mode. To investigate the effects of different operating modes on the SLAM performance in 2D, I have used 13 sensors in even layout on the horizontal plane and introduced each sensor parameter one by one.

I have found that by increasing the sensor resolution to 3x3 or 4x4, the SLAM performance is increased and mapping works reliably. Sensor resolutions 1x1 and 2x2 have significantly higher update rate, but using these settings, the SLAM algorithm became unreliable.

As next I have investigated the effects of sampling rate and reduced it for resolutions 3x3 and 4x4 accordingly. After proper tuning of Cartographer, the accuracy of the two resolutions proved to be similar, while lower resolution provides more frequent pose estimates. Limiting the maximum distance of the two setups had a significant effect of the performance and Cartographer couldn't track the drone's position in wide open spaces anymore.

After simulating all parameters of the sensors, I have decreased the number of LIDARS while keeping an even layout. The tracking proved to be accurate using at least 8 sensors in total.

The evaluation of 2D SLAM performance helped the process of 3D evaluation. In this case I have also chose an even layout and simulated 43 LIDARs in total and introduced each sensor parameter one by one. 3D mapping worked both for 3x3 and 4x4 sensor resolutions reliable. Even with lowered sampling rate according to each resolution, the mapping quality proved to be adequate. At last I have limited the maximum range to 2.75m in both cases. The ranging limit caused the SLAM algorithm to be unreliable and produced maps became messy. The fact that 3D SLAM is unreliable even with 43 sensors, proves that VL53L1X sensors and Cartographer SLAM cannot be used for 3D SLAM. Using higher number of sensors might increase the performance, but placing 43 sensors on a real-world drone is already unfeasible.

At last I have tested the performance of 2D SLAM using 8 sensors, by mapping a different building that has rooms of varying size. The algorithm performed well in small to mid sized rooms, but in open spaces it completely lost track and was not able to compensate

this error. I have found that the system works reliably, when walls on all sides are in range of all the sensors. Therefore the maximum room size is 2.75x2.75m, where 2D SLAM using 8 VL53L1X sensors work reliably.