

POLITECNICO DI MILANO

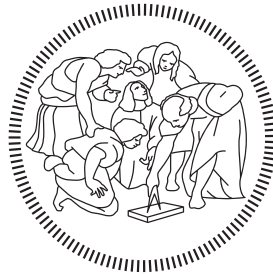
Facoltà di Ingegneria

Scuola di Ingegneria Industriale e dell'Informazione

Dipartimento di Elettronica, Informazione e Bioingegneria

Master of Science in

Computer Science and Engineering



## A ROS implementation of a 6-DoF EKF for indoor drone Visual SLAM

Advisor: PROF. MATTEO MATTEUCCI

Master Graduation Thesis by:

DIEGO EMANUEL AVILA

Student Id n. 903988

Academic Year 2019-2020



*My dedication*



---

## ACKNOWLEDGMENTS

---



---

## CONTENTS

---

Abstract	xi
<b>1 BACKGROUND</b>	<b>1</b>
1.1 Robot Operating System	1
1.1.1 Nodes	1
1.1.2 Topics	1
1.1.3 Services	2
1.1.4 Tools	2
1.1.5 MAVLink and MAVROS	3
1.1.6 RTAB-Map	3
1.2 Transformations	5
1.2.1 Translation	5
1.2.2 Rotation	6
1.2.3 Homogeneous Transform	7
1.3 Kalman Filter	8
1.3.1 Extended Kalman Filter	10
1.4 Simultaneous Localization and Mapping	12
1.4.1 EKF-SLAM	15
1.4.2 Adding new landmarks	17
1.5 Normalized Estimation Error Squared	18
<b>2 EKF-SLAM IMPLEMENTATION</b>	<b>21</b>
2.1 The Drone	21
2.1.1 Characteristics	21
2.1.2 Reference Frames	22
2.2 Motion Model	24
2.3 Observation Models	27
2.3.1 Observation model for laser range	28
2.3.2 Observation model for Poles	28
2.3.3 Observation model for Markers	30
2.4 Visual process pipeline	32
2.4.1 Visual process for Poles	33
2.4.2 Visual process for Markers	34
2.5 Overall Architecture	34
2.5.1 ROS nodes	36
<b>3 EXPERIMENTAL RESULTS</b>	<b>39</b>
3.1 The Environment	39
3.2 Simulated Experiments	39
3.2.1 Procedures	39
3.2.2 Results	39

4	CONCLUSIONS AND FUTURE WORK	41
4.1	Conclusion . . . . .	41
4.2	Future Work . . . . .	41
	BIBLIOGRAPHY	43
A	DOCUMENTATION	45
B	USER MANUAL	47



---

## LIST OF FIGURES

---

Figure 1.1	ROS basic concept . . . . .	2
Figure 1.2	Octomap example . . . . .	4
Figure 1.3	RTAB-Map example . . . . .	5
Figure 1.4	Rotation around origin . . . . .	6
Figure 1.5	Rotation around axis $z$ followed by rotation about axis $x$ . . . . .	8
Figure 1.6	Example of a Homogeneous transform . . . . .	9
Figure 1.7	Linear and nonlinear transformation of a Gaussian random variable . . . . .	11
Figure 1.8	Linearization applied in EKF . . . . .	12
Figure 1.9	Example of SLAM problem . . . . .	14
Figure 2.1	3DR Iris Quadrotor frame. . . . .	21
Figure 2.2	Main reference frames in the system. . . . .	24
Figure 2.3	NED to ENU conversion scheme. [9] . . . . .	25
Figure 2.4	Range and Bearing example . . . . .	29
Figure 2.5	Example of the drone observing a marker . . . . .	30
Figure 2.6	Components diagram of the EKF Localization node . . . . .	34
Figure 2.7	Class diagram of the EKFSLAM node . . . . .	35
Figure 2.8	Detail of the EKF-SLAM node interactions . . . . .	37
Figure 2.9	ROS graph of the system . . . . .	38
Figure 3.1	Gazebo simulated environment. . . . .	40

---

## LIST OF TABLES

---

Table 2.1	PixHawk 4 Specification . . . . .	23
-----------	-----------------------------------	----

---

## ACRONYMS

---

IDL	Interface Description Language
-----	--------------------------------

<b>ROS</b>	Robot Operating System
<b>RViz</b>	ROS Visualization
<b>SLAM</b>	Simultaneous Localization And Mapping
<b>EKF</b>	Extended Kalman Filter
<b>KF</b>	Kalman Filter
<b>NEES</b>	Normalized Estimation Error Squared
<b>ENU</b>	East-North-Up
<b>NED</b>	North-East-Down
<b>RTAB-Map</b>	Real-Time Appearance-Based Mapping
<b>ViSP</b>	Visual Servoing Platform

---

## ABSTRACT

---

---

## SOMMARIO

---

---

## BACKGROUND

---

### 1.1 ROBOT OPERATING SYSTEM

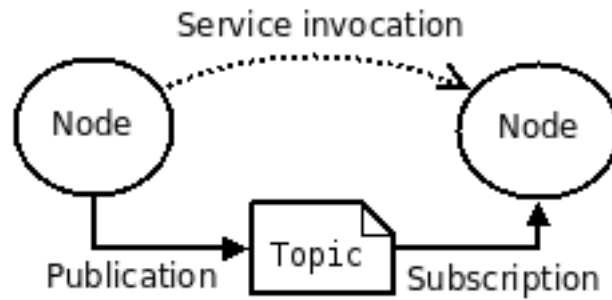
The Robot Operating System ([ROS](#)) is an open source middleware that provides inter-process communication through a message passing mechanism. It is also a collection of libraries, tools and conventions with the aim of simplify the development of software for robots. In this sense, one of the main components of the framework are the nodes, which encapsulate processes and/or algorithms.

#### 1.1.1 *Nodes*

Nodes are processes that perform a specific computation. In the system the nodes communicate between them using a publisher/subscriber infrastructure based on topics. A node subscribes to a particular topic, and it will receive messages from a publisher node. This way, a publisher node is hid to the subscriber, reducing the coupling between them. The advantage of this mechanism is that nodes can be developed separately once the message structure is defined, forcing developers to implement clear interfaces for communication by using a message Interface Description Language ([IDL](#)). Moreover, this enables a modular and distributed development of the robotic system, while providing some fault tolerance and reduced code complexity.

#### 1.1.2 *Topics*

As mentioned before, communication between nodes is done via a publisher/-subscriber architecture based on topics, which are named buses over which messages are exchanged. There can be several publishers (as well as subscribers) for a topic. A node that generates data publishes all its information in a specific topic bus, consequently this information is consumed by those nodes subscribed to that topic. The transport protocol used for exchanging messages is defined at runtime and can be of two types: TCPROS or UDPROS.



**Figure 1.1:** ROS basic concepts. Two nodes communicate between them using the messages published in a topic. One node is responsible of publishing the topic, while the other subscribes and receives the messages of that particular topic. Furthermore, the node in the left calls a service provided by the node in the right, so that node will perform some computation and will return the result. [1]

The description of these protocols is beyond the scope of this document and the reader is invited to read the ROS documentation for detailed information.

### 1.1.3 Services

As well as topics, services are a way to communicate nodes. The difference is that topics are an asynchronous way of communication, while services are synchronous. The key difference lies on the ability of nodes to decide when to trigger a service. Hence, services are used in between nodes to retrieve specific information or to request another node to carry out a computation beyond caller's node scope.

### 1.1.4 Tools

Regarding the tools provided by ROS, one that is crucial for debugging and/or experimentation is the bag recording and playback. Since the exchange of messages between nodes is anonymous (meaning that nodes communicate between each others without knowing which node sent or received a message) it is possible to record the messages during a period of time, without taking into consideration which node sent a message and which node received it. This recorder bag is useful for debugging since it can be played back, hence reproducing a previous experiment. Also, it can be used for the development of new nodes that depend on the messages contained in the bag.

Another useful tool provided is ROS Visualization ([RViz](#)), a 3D visualization tool. With this tool it is possible to see the robot, orientation, reference frames, covariance matrices, etc. In addition, it is possible to draw lines, arrows, text

and others, onto the environment in order to see useful information that cannot be extracted from messages.

#### 1.1.5 *MAVLink and MAVROS*

MAVLink is a binary telemetry protocol designed for resource-constrained systems and bandwidth-constraint links, more specifically for drones of all kinds. As with ROS, it adopts a publisher-subscriber architecture, where data streams are published as topics. Its key features, as published in their website are:

- Since its messages do not require any special framing, it is well suited for applications with very limited bandwidth.
- It provides methods for detecting package drops, corruption and for package authentication.
- Allows up to 255 concurrent systems on the network.
- Enables both offboard and onboard communications.

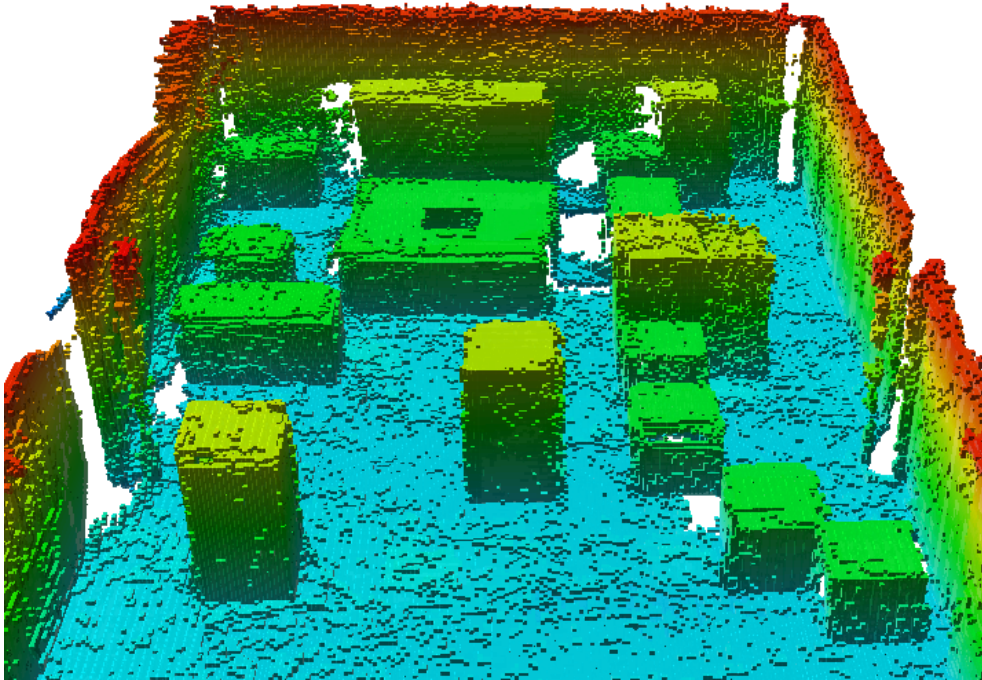
MAVROS is the extension of MAVLink in ROS, with the additive of being a proxy for Ground Control Station tool. Its main features, as published in its website are:

- Communication with autopilot via serial port, UDP or TCP .
- Internal proxy for Ground Control Station (serial, UDP, TCP).
- Plugin system for ROS-MAVLink translation.
- Parameter manipulation tool.
- Waypoint manipulation tool.
- PX4Flow support (by mavros\_extras).
- OFFBOARD mode support.

One characteristic that worth mentioning, is that it translates the North-East-Down ([NED](#)) reference frames into East-North-Up ([ENU](#)) reference frames, and vice-versa, so as to be compliant with ROS standard reference frames.

#### 1.1.6 *RTAB-Map*

Real-Time Appearance-Based Mapping ([RTAB-Map](#)) is a RGB-D, stereo and LiDAR graph-based Simultaneous Localization And Mapping ([SLAM](#)) approach based on an incremental appearance-based loop closure detector.



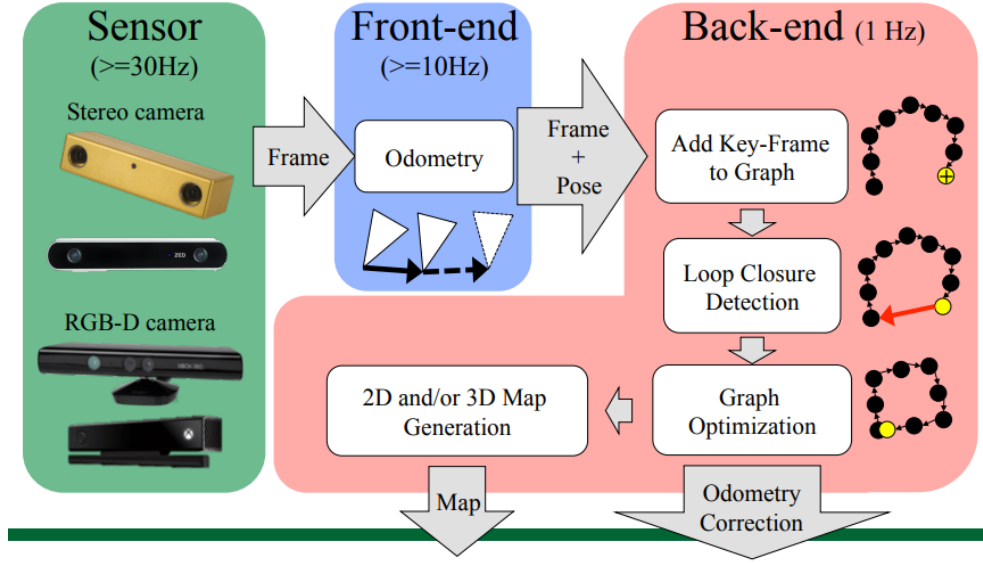
**Figure 1.2:** Octomap example. Octotree representation generated from data, showing occupied voxels only. [14]

It allows to build a 3 dimensional map of the environment using a stereo camera, a LiDAR sensor and/or the odometry estimation. The map is built using a 3D occupancy grid by means of Octomap library, which builds a tree-based representation of the mapped area. Octomap library “performs a probabilistic occupancy estimation to ensure updatability and to cope with sensor noise”[14].

The Octomap library generates an octotree, which is a hierarchical data structure used to generate spacial subdivision in 3 dimensions. Each node in an octotree represents a voxel, which is a cubic volume, which is at the same time, subdivided in eight smaller cubes until a given minimum voxel size called resolution. This way, the octotree data structure is used to build an occupancy grid of a volume. An example of a octotree can be seen in Figure 1.2, where the color of each occupied voxel represents its height: higher voxels are red-colored, while lower voxels are light-blue-colored.

RTAB-Map library, uses the Octomap library to store the 3D occupancy grid of the environment. The `rtabmap_ros` node is the extension of the RTAB-Map library for ROS. It uses the information of the stereo camera or a RGB-D camera and/or a LiDAR information to create a cloud point that is used to





**Figure 1.3:** RTAB-Map example. An stereo camera or a RGB-D camera is used among the odometry estimation. Both are used to generate a 3D occupancy grid using Octomap library, and to correct the odometry estimation. [15]

build the map. Furthermore, it uses the provided odometry to estimate and correct the robot's pose into the 3D map. An example of this procedure can be seen in Figure 1.3.

## 1.2 TRANSFORMATIONS

A robot, in this case the drone, can be considered as a rigid body that moves and rotates around the environment. Hence, it is necessary to model the drone displacement in space, and this can be done by decomposing the movement as translations and rotations.

### 1.2.1 Translation

Given a vector  $\mathbf{b} = \begin{bmatrix} b_x & b_y & b_z \end{bmatrix}^T$  that represents the center of mass of a rigid body, a translation that displaces  $\mathbf{b}$  parallel to itself of a given vector  $\mathbf{t}$  can be defined as:

$$\mathbf{b} + \mathbf{t} = \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} b_x + t_1 \\ b_y + t_2 \\ b_z + t_3 \end{bmatrix}$$

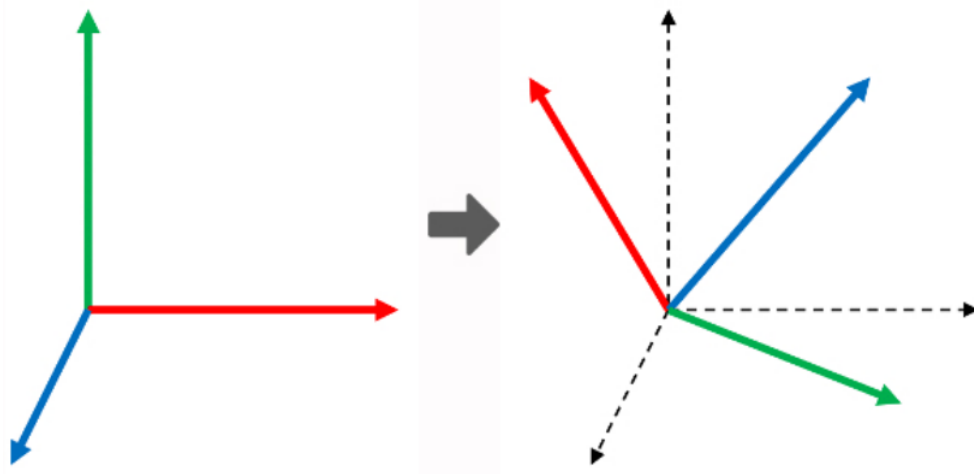


Figure 1.4: cambiar esto

Given a point  $\mathcal{P}$  represented in the reference frame  $\mathcal{R}_m$  by  $v_P^m$  will be represented in the reference frame  $\mathcal{R}_n$  by:

$$v_P^n = v_P^m + t_m^k$$

### 1.2.2 Rotation

In three dimensional space, any displacement of a rigid body is equivalent to a single rotation of a given angle about some axis that contains the point. It is possible to express the previous statement as  $v = u * \theta$ , being  $u$  a unit vector representing the axis and  $\theta$  the angle.

Unlike with translations, two or more rotations cannot be simply the sum of the related vectors. A rotation can be described assuming a rigid body in a reference frame with its origin fixed, while the unit vectors are changed under the rotation. This way, a rotation is characterized by the mathematical relation of these two reference frames. Hence, in order to represent a rotation, two reference frames with a common origin are needed, as shown in Figure 1.4. At the beginning the two reference frames coincide, then one of them is rotated around the origin by an arbitrary angle. After this procedure, the two reference frames are not coincident anymore, however, both share the same origin.

The rotation by any angle of any axis can be represented in a matrix form,  $R_n^m$ , where each column of the matrix represents the unit vector of  $\mathcal{R}_n$  in  $\mathcal{R}_m$ , so it represents the rotated frame  $\mathcal{R}_n$  with respect the fixed frame  $\mathcal{R}_m$  with common origin. Furthermore, it is possible to show that matrix  $R_n^m$  is

*orthonormal*, which means that its inverse is equal to its transpose.

Given the matrix representation of a rotation, it is possible to provide a way to rotate a vector in the space. When the reference frames origins are the same, the rotation of a vector is possible by multiplying it by the rotation matrix around an axis:  $\mathbf{v}^m = \mathbf{R}_n^m \mathbf{v}^n$ . This kind of rotations are called *elementary rotations*, and given that there are three axis in a 3D space, three different elementary rotations can be defined:

1. Rotation around axis x:

$$\mathbf{R}_{x,\theta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

2. Rotation around axis y:

$$\mathbf{R}_{y,\theta} = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

3. Rotation around axis z:

$$\mathbf{R}_{z,\theta} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

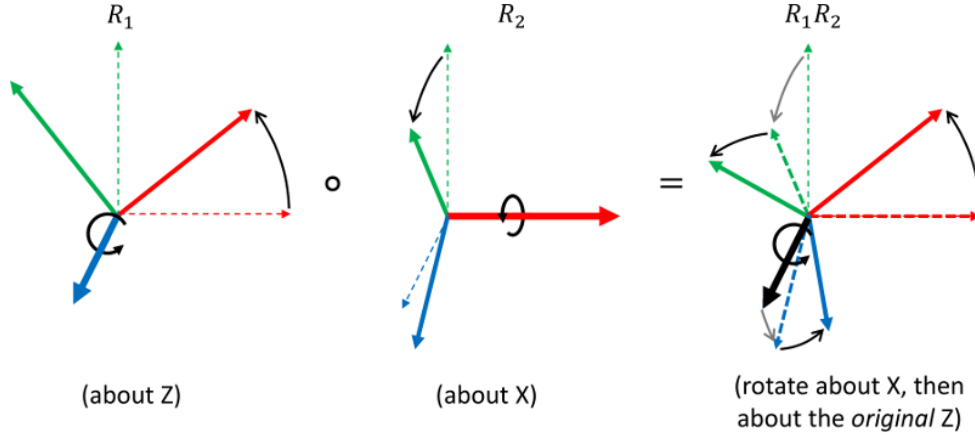
As mentioned, unlike translations, rotations cannot be summed. Hence, if two consecutive rotations need to be performed around two axis, a multiplication is needed:

$$\mathbf{R}_{y,x} = \mathbf{R}_{x,\alpha} * \mathbf{R}_{y,\beta} \quad (1.1)$$

where  $\mathbf{R}_{y,x}$  is the rotation around axis y by an angle  $\beta$  followed by a rotation around the new axis x by an angle  $\alpha$ , as shown in Figure 1.5.

### 1.2.3 Homogeneous Transform

A rotation and a translation can be accomplished as follow:  $\hat{\mathbf{v}} = \mathbf{R}\mathbf{v} + \mathbf{t}$ . Nevertheless, this can be achieve by using *homogeneous coordinates*, where the point  $\mathbf{v}$ , expressed in its homogeneous representation, is  $\mathbf{v} = \begin{bmatrix} v_x & v_y & v_z & 1 \end{bmatrix}^T$ .



**Figure 1.5:** Rotation about axis  $z$  followed by rotation about axis  $x$ . The second rotation is performed about the original axis  $z$ . This final rotation can be achieved by multiplying  $R_1$  and  $R_2$ . [16]

This way, the rotation and translation of a point can be accomplished by multiplying the 3D point with the homogeneous transformation matrix:

$$T_m^n = \begin{bmatrix} R_m^n & t \\ \mathbf{o} & 1 \end{bmatrix} \quad (1.2)$$

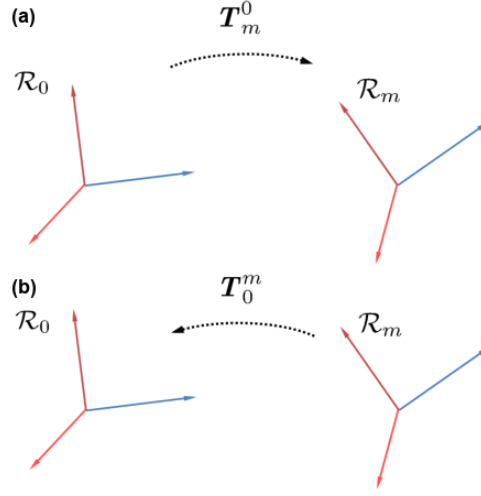
$$\hat{v}^n = T_m^n v^m \quad (1.3)$$

As can be noticed in equation (1.2), the homogeneous transformation is represented by a  $4 \times 4$  matrix, and unlike the pure rotation matrix showed in Section 1.2.2, the inverse of the homogeneous transformation is not equal to its transpose. The inverse of the homogeneous transformation is equivalent to:

$$T^{-1} = \begin{bmatrix} R^T & -R^T t \\ \mathbf{o} & 1 \end{bmatrix} \quad (1.4)$$

### 1.3 KALMAN FILTER

Introduced by [4], the Kalman Filter (KF) provides a recursive method for estimating the state of a dynamic system with presence of noise. It is a technique for filtering and prediction in *linear Gaussian systems* and applicable only to continuous states. One of its key features is that it simultaneously maintains estimates of the state vector and the estimate error covariance. It assures that the posteriors are Gaussian if the Markov assumption is hold, in addition to three conditions. The Markov assumption states that the past and future data are independent if one knows the current state; the other three conditions are the following:



**Figure 1.6:** Example of a Homogeneous transform. (a) Transformation  $T_m^0$  is applied to a point in reference frame  $\mathcal{R}_0$  to express it in reference frame  $\mathcal{R}_m$ . (b) The inverse transformation is applied, so to express the point in reference frame  $\mathcal{R}_m$  into reference frame  $\mathcal{R}_0$ . [12]

1. The state transition probability  $p(x_t | u_t, x_{t-1})$  must be a linear function in its arguments. This is guaranteed by  $x_t = A_t x_{t-1} + B_t u_t + \epsilon_t$ . Here,  $x_t$  and  $x_{t-1}$  are the state vectors, of size  $n$ , and  $u_t$  represents the control vector, of size  $m$ , at time  $t$ ;  $A_t$  and  $B_t$  are matrices of size  $n \times n$  and  $n \times m$  respectively. In this way, the state transition function becomes linear in its arguments, hence KF assumes a linear system dynamics.

$\epsilon_t$  represents the uncertainty introduced by the state transition, and it is a Gaussian random variable with zero mean and  $R_t$  covariance.

2. The measurement probability  $p(z_t | x_t)$  must be linear in its arguments. This is guaranteed by  $z_t = C_t x_t + \delta_t$ , where  $z_t$  represents the measurement vector of size  $k$ ,  $C_t$  is a matrix of size  $k \times n$  and  $\delta_t$  is a Gaussian noise with zero mean and  $Q_t$  covariance.
3. The initial belief should be normally distributed.

Given these three conditions, we are guaranteed that the posterior probability is Gaussian.

The KF can be seen in Algorithm 1.1 and its input is the belief at time  $t - 1$ , represented by its mean ( $\mu_{t-1}$ ) and its covariance ( $\Sigma_{t-1}$ ), in addition to the control vector ( $u_t$ ) and the observations ( $z_t$ ). As a result, it returns the current belief characterized by its mean  $\mu_t$  and covariance  $\Sigma_t$ .

The first step in the algorithm (lines 1 and 2), represent the prediction step. It calculates the current belief before incorporating the observations, but after

---

**Algorithm 1.1:** Kalman Filter algorithm

---

**Input:**  $\mu_{t-1}, \Sigma_{t-1}, u_t, z_t$

- 1  $\hat{\mu}_t = A_t \mu_{t-1} + B_t u_t$
- 2  $\hat{\Sigma}_t = A_t \Sigma_{t-1} A_t^T + R_t$
- 3  $K_t = \hat{\Sigma}_t C_t^T (C_t \hat{\Sigma}_t C_t^T + Q_t)^{-1}$
- 4  $\mu_t = \hat{\mu}_t + K_t (z_t - C_t \hat{\mu}_t)$
- 5  $\Sigma_t = (I - K_t C_t) \hat{\Sigma}_t$
- 6 **return**  $\mu_t, \Sigma_t$

---

adding the control vector. The estimated belief is characterized by its mean,  $\hat{\mu}_t$ , and its covariance  $\hat{\Sigma}_t$ .

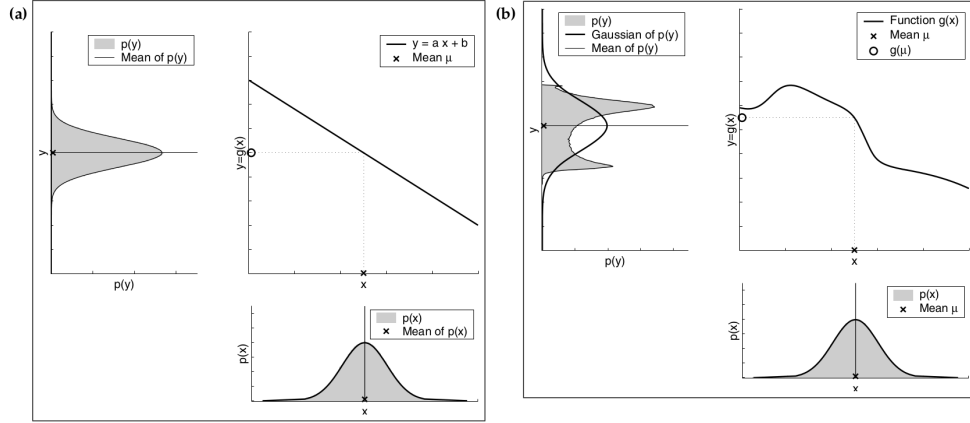
The second step (from line 3 to 5) starts by calculating  $K_t$ , the Kalman gain, which specifies the degree in which the observation is incorporated to the new state estimate.  $K_t$  can be seen as the weighting factor that weights the relationship between the accuracy of the predicted state estimate and the observation noise. When  $K_t$  is large, the observations have more importance in the final estimate; while if  $K_t$  is small, the observations do not have much importance in the correction step. Following, at line 4, the new mean is estimated by means of the Kalman gain and the *innovation*, which is the difference between the observation  $z_t$  and the expected measurement  $C_t \hat{\mu}_t$ . Finally, the new covariance is calculated.

### 1.3.1 Extended Kalman Filter

The linearity conditions that make the KF to work are, in some cases, far from reality: state transition functions and measurements are rarely linear in practice. The Extended Kalman Filter ([EKF](#)) works through a process of linearization, where nonlinear state transition and observation functions are approximated by a Taylor series expansion.

The Figure [1.7a](#) shows the linear transformation of a random Gaussian variable, whose density function is  $\mathcal{N}(x; \mu, \sigma^2)$ . Assuming that the random variable is transformed using a linear function  $y = ax + b$ , the resulting random variable will be Gaussian with mean  $a\mu + b$  and variance  $a^2 \sigma^2$ .

However, as shown in Figure [1.7b](#), this does not happen if the transformation is not linear. In this case, assuming the original random variable is transformed using a nonlinear function  $g$ , the density of the resulting random



**Figure 1.7:** Linear (a) and nonlinear (b) transformation of a Gaussian random variable. The lower right plot shows the density function of the random variable. The upper right plot shows the transformation of the random variable. The upper left plot shows the resulting density function. [2]

variable is not Gaussian anymore.

The state transition probability and observation probabilities are ruled by nonlinear functions  $g$  and  $h$  respectively. Matrices  $A$  and  $B$  are replaced by function  $g(u_t, x_{t-1})$  and matrix  $H$  is replaced by function  $h(x_t)$ , making the belief not Gaussian. This is solved in EKF by approximating to the true belief, not the exact one as happens with linear KF. The approximation is done using a linearization method that approximates the nonlinear function by a linear function that is tangent to it, thereby maintaining the Gaussian properties of the posterior belief.

The used method is the first order Taylor expansion, which constructs a linear approximation of a function  $g$  from  $g$ 's value and slope, which is given by

$$g'(u_t, x_{t-1}) = \frac{\partial g(u_t, x_{t-1})}{\partial x_{t-1}} \quad (1.5)$$

Since  $g$  depends on the control variable  $u$  and the state  $x$ , we need to define a value for  $x$ , and the logical choice is the mean of the posterior in the previous time step:  $\mu_{t-1}$ . This way

$$g(u_t, x_{t-1}) = g(u_t, \mu_{t-1}) + g'(u_t, \mu_{t-1})(x_{t-1} - \mu_{t-1}) \quad (1.6)$$

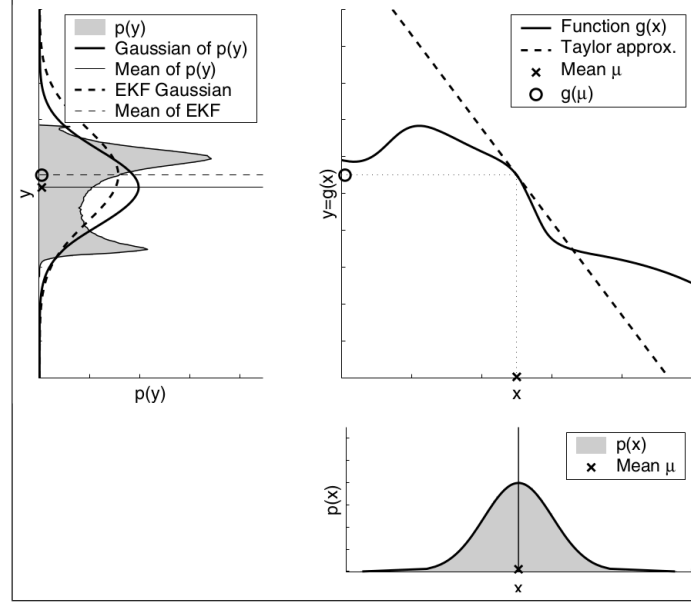
where  $g'$  is the *Jacobian* of  $g$ , usually expressed as  $G_t$ , and it depends on  $u_t$  and  $\mu_{t-1}$ , hence it changes through time.

The same linearization is applied to the observation function  $h$ :

$$h(x_t) = h(\hat{\mu}_t) + h'(\hat{\mu}_t)(x_t - \hat{\mu}_t) \quad (1.7)$$

$$h'(\hat{\mu}_t) = \frac{\partial h(x_t)}{\partial x_t} \quad (1.8)$$

where  $h'$  is the Jacobian of  $h$ , usually expressed as  $H_t$ . In this case, the linearization is done around  $\hat{\mu}_t$ , which is the state estimate just before computing  $h$ .



**Figure 1.8:** Linearization applied in EKF. In this case, the nonlinear function  $g$  is approximated using first order Taylor expansion, that is a linear function tangent to  $g$  at the mean of the original density function. The linearization is not perfect, so it adds an error, depicted in the upper left plot. This error is the difference between the dashed line and the solid line. [2]

The Figure 1.8 depicts the approximation of  $g$  by a linear function that is tangent around its mean. The resulting density function is shown in the upper left plot with a dashed line, that is similar to the original density function.

The EKF algorithm can be seen in Algorithm 1.2, and it is similar to Algorithm 1.1. The difference lies in the usage of the nonlinear functions  $g$  and  $h$  and their Jacobians,  $G_t$  and  $H_t$  respectively.

#### 1.4 SIMULTANEOUS LOCALIZATION AND MAPPING

Among all the problems faced by autonomous mobile robots, two of them are relevant for this work: localization and mapping. The former one, is related to the problem of where the robot is, while the later is related to building a map of the environment. However, in order to accurately localize itself the



---

**Algorithm 1.2:** Extended Kalman Filter algorithm

---

**Input:**  $\mu_{t-1}, \Sigma_{t-1}, u_t, z_t$

- 1  $\hat{\mu}_t = g(u_t, \mu_{t-1})$
- 2  $\hat{\Sigma}_t = G_t \Sigma_{t-1} G_t^T + R_t$
- 3  $K_t = \hat{\Sigma}_t H_t^T (H_t \hat{\Sigma}_t H_t^T + Q_t)^{-1}$
- 4  $\mu_t = \hat{\mu}_t + K_t (z_t - h(\hat{\mu}_t))$
- 5  $\Sigma_t = (I - K_t H_t) \hat{\Sigma}_t$
- 6 **return**  $\mu_t, \Sigma_t$

---

robot needs a map of the environment in which it is immersed in, and, in order to build a map, it needs to know where it currently is, giving us a chicken-egg situation. Hence, the [SLAM](#) problem appears when the robot has no knowledge of its localization nor its environment map, while measurements and controls are given.

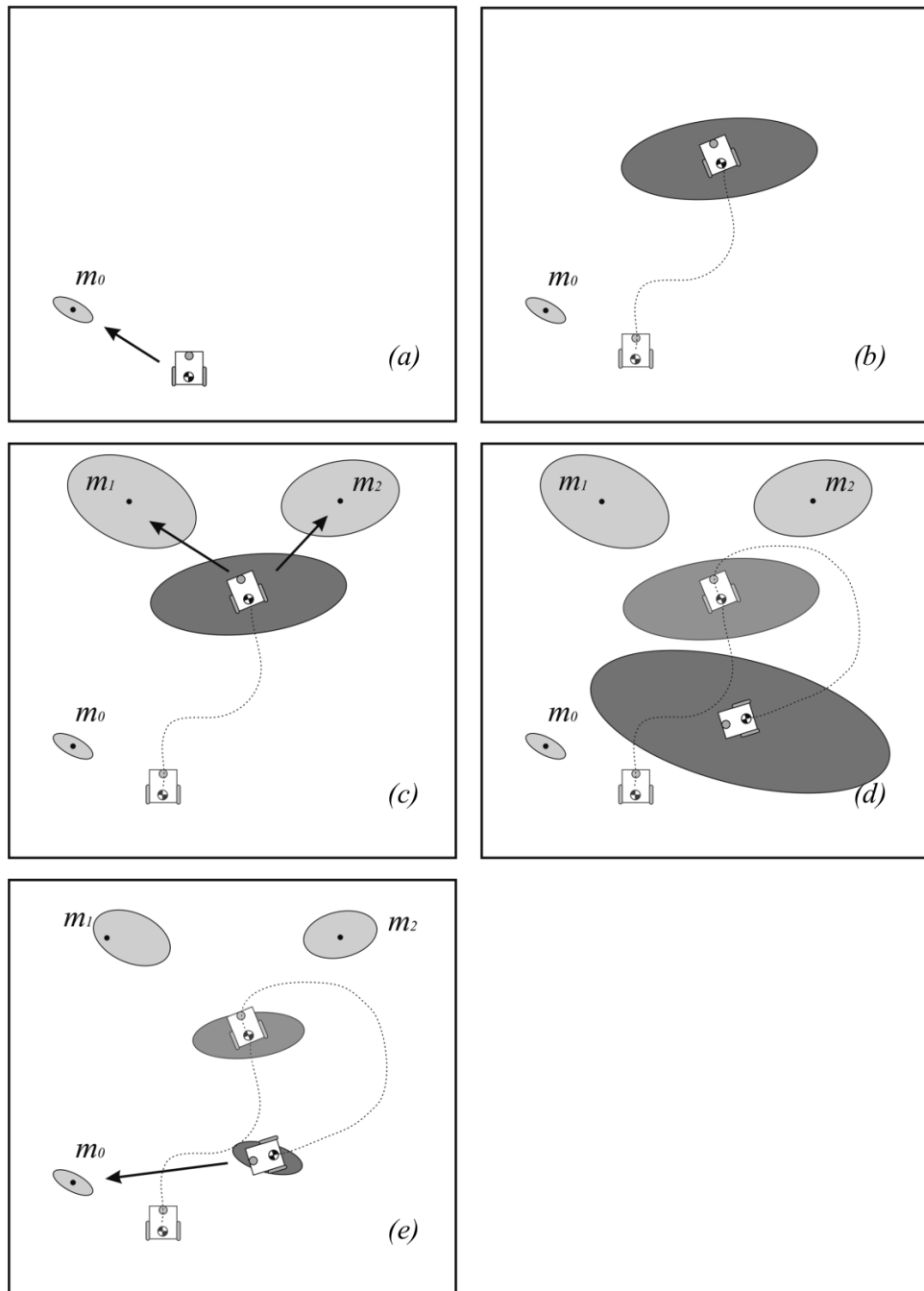
The problem of building a map can be summarized in the following steps:

1. The robot senses the environment using its sensors
2. It creates a representation of the acquired data
3. It integrates the processed sensor data with the previously learned map structure

While this process can be done by manually moving the robot around the environment, it is more challenging to build the map while the robot is moving autonomously.

On the other hand, assuming that the robot already knows the map, the localization problem could be trivial if no noise is present at all. The sensors, wheel encoders, different kinds of terrain, battery life, etc, all of these can make the robot to increase its uncertainty related to where it is. As robot moves around the environment it uses its sensors to estimate its position, increasing its uncertainty regarding its position relative to the map. At some point, the robot will "see" a known landmark or feature in the environment, correcting its position while reducing the uncertainty.

The localization and mapping problems can be solved together by using a SLAM technique, with which the robot will build the map while localizing itself in it. An example of this problem can be seen in [Figure 1.9](#), where a robot moves around the environment and sees some features or landmarks. The uncertainty regarding its position is low when it starts, and keeps growing



**Figure 1.9:** At the beginning (a) the robot has low uncertainty regarding its pose. As it moves around the environment its uncertainty, represented by the dark gray ellipsis, grows (b), (c), (d), until it sees a known landmark (e), making the position uncertainty to shrink. [5]

while it moves around. At the end, it sees a known landmark ( $m_0$ ) making the uncertainty to shrink. As can be seen in the figure, the robot adds new landmarks to the map ( $m_1$  and  $m_2$ ) with their corresponding uncertainty, and when the robot sees the first landmark, not only its own uncertainty decreases, but also the two new landmarks' uncertainty. In this way, the robot's position is correlated with the observations' position estimates.

The idea of the SLAM problem is to estimate a posterior belief that involves not only the robot pose, but also the map:  $p(x_t, m|z_{1:t}, u_{1:t})$ , where  $x_t$  is the robot's pose at time  $t$ ,  $m$  is the map,  $z_{1:t}$  are the measurements, and  $u_{1:t}$  are the controls given to the robot.

#### 1.4.1 EKF-SLAM

The SLAM problem can be addressed, between others, using an EKF approach. The algorithm proceeds in the same way as shown in Section 1.3.1, being  $\mu$  a state vector containing the information for the robot pose ( $q_r$ ) and the landmarks' pose ( $m_i$ ):

$$\mu = \begin{bmatrix} q_r & m_0 & \dots & m_{n-1} \end{bmatrix}^T \quad (1.9)$$

One thing that is worth mentioning is that in EKF-SLAM maps are *feature based*. This means that the features or landmarks are assumed to be points in the space: if the robot sees, for example, a chair, it will store a point that represents the chair in space. Also, as explained before, it assumes Gaussian noise for the robot motion and observations.

The EKF-SLAM algorithm estimates the robot's pose in addition to all encountered landmarks' poses along its way. Thus, there is a correspondence between robot's pose and landmarks, and that is why it is necessary to include the landmarks information into the state vector. Hence, the algorithm estimates the posterior  $p(\mu_t|z_t, u_t)$ .

Assuming the robot's pose is composed by  $x, y, \theta$ , the markers' pose is composed by  $x, y$ , and there are  $N$  markers, the state vector  $\mu$  will have a length of  $3 \times 2N$ , and the covariance matrix  $\Sigma$  will have a size of  $(3 \times 2N) \times (3 \times 2N)$ .

The algorithm can be seen in Algorithm 1.3. From lines 1 to 3, it computes the state vector and covariance matrix updates; the function `computeJacobian` computes, indeed, the Jacobian matrix for the motion model  $g$ , and the result-

**Algorithm 1.3:** EKF-SLAM algorithm

---

**Input:**  $\mu_{t-1}, \Sigma_{t-1}, u_t, z_t$

```

1  $\hat{\mu}_t = g(\mu_{t-1}, u_t);$ 
2  $G_t = \text{computeJacobian}(g);$ 
3  $\hat{\Sigma}_t = G_t \Sigma_{t-1} G_t^T + R_t;$ 
4 foreach landmark observation  $z_t^i$  do
5   if landmark  $i$  has not being seen before then
6      $\text{addLandmarkToStateVector}(z_t^i)$ 
7    $H_t^i = \text{computeJacobian}(h^i);$ 
8    $v^i = z_t^i - h^i(\hat{\mu}_t);$ 
9    $S = H_t^i \hat{\Sigma}_t H_t^{iT} + Q_t;$ 
10   $K_t^i = \hat{\Sigma}_t H_t^{iT} S^{-1};$ 
11   $\mu_t = \hat{\mu}_t + K_t^i (v^i);$ 
12   $\Sigma_t = (I - K_t^i H_t^i) \hat{\Sigma}_t;$ 
13 return  $\mu_t, \Sigma_t$ 

```

---

ing matrix has the same size as the covariance matrix and has the following characteristic:

$$G_t = \begin{bmatrix} G_r & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix} \quad (1.10)$$

$$G_r = \begin{bmatrix} \frac{\partial x'}{\partial \mu_{t-1,x}} & \frac{\partial x'}{\partial \mu_{t-1,y}} & \frac{\partial x'}{\partial \mu_{t-1,\theta}} \\ \frac{\partial y'}{\partial \mu_{t-1,x}} & \frac{\partial y'}{\partial \mu_{t-1,y}} & \frac{\partial y'}{\partial \mu_{t-1,\theta}} \\ \frac{\partial \theta'}{\partial \mu_{t-1,x}} & \frac{\partial \theta'}{\partial \mu_{t-1,y}} & \frac{\partial \theta'}{\partial \mu_{t-1,\theta}} \end{bmatrix} \quad (1.11)$$

where  $\frac{\partial x'}{\partial \mu_{t-1,x}}$  is the derivative of  $g$  along  $x'$  dimension, taken with respect to  $x, y, \dots$  at  $\mu_{t-1}$ .

At line 4, it iterates through every observation  $z_t$ . At each time step  $t$ , a sensor obtains a set of observations  $z^i$  of one of the  $N$  landmarks. Each observation is associated with a map feature, and this association is accomplished by a prediction of the measurement that each feature would generate and a measure of the difference between the prediction and the sensor measurement. The prediction of the measurements can be obtained by the observation model  $h^i$ , which result is a vector of predicted features. If the observation  $z^i$  comes from a landmark  $i$ , the following relation is hold:

$$z^i = h^i(x_t) + w^i \quad (1.12)$$

where  $x_t$  is the true state and  $w^i$  is the observation noise with covariance  $Q_t$ , and assumed to be zero mean, Gaussian, additive and independent of the process noise. If the landmark is not already in the state vector, at line 6 it is added by projecting the observation and calculating the landmark's pose, adding two new elements to the state vector and two new more columns and rows to the covariance matrix. At line 7 the Jacobian of the observation model is computed.

At line 8, the innovation is calculated while its covariance is calculated at line 9. The innovation measures the discrepancy between the predicted observation and the actual sensor measurement. At line 10 the Kalman gain is computed, and at line 11 the state vector is updated. The gain propagates the information through all the state vector, updating not only the robot's pose, but also the landmarks' poses.

The fact that the Kalman gain is not sparse is important, because observing a landmark not only improves the estimate of that landmark's pose, but also all the others, along with the robot's pose. This effect can be seen in Figure 1.9, with an additional explanation: most of the uncertainty of the landmarks' poses is caused by the robot's own uncertainty, so the location of those previously seen landmarks are correlated. When the robot gains information about its own pose, this information is propagated to the landmarks, and as result it improves the localization of other landmarks in the map.

#### 1.4.2 Adding new landmarks

So far, the existence and accuracy of a map was assumed. Nevertheless, this is not always true, and the construction of a map should somehow be done. Hopefully, EKF-SLAM algorithm considers the possibility of adding new landmarks to the map while doing the localization process with known landmarks. This process of adding new landmarks to the map is done via what is called inverse observation model, which handles the sensor measurements, identifies the new landmark, and adds it to the state vector.

The inverse observation model will produce the coordinates of the new landmark in the map, and this information will be added to the state vector, which will increase its size (in this case) by two new elements.

$$h^{-1}(x_r, z) = \begin{bmatrix} x_l \\ y_l \end{bmatrix} \quad (1.13)$$

$$x_t = y(x, z) = \begin{bmatrix} x \\ h^{-1}(x_r, z) \end{bmatrix} \quad (1.14)$$

In equation (1.13)  $x_r$  refers to the elements in the state vector that corresponds to the robot's pose, and  $z$  refers to the observation elements, which can be, for example, range and bearing data. Since the state vector has a variable length, extending the covariance matrix is also needed whenever the robot sees a landmark that has not being seen previously. The extension of the covariance matrix is achieved as shown in equation (1.15).

$$\hat{\Sigma} = Y_x \Sigma Y_x^T + Y_z Q_t Y_z^T \quad (1.15)$$

$$Y_x = \frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial x}{\partial x} \\ \frac{\partial h^{-1}}{\partial x} \end{bmatrix} = \begin{bmatrix} I \\ G_x \end{bmatrix} \quad (1.16)$$

$$Y_z = \frac{\partial y}{\partial z} = \begin{bmatrix} \frac{\partial x}{\partial z} \\ \frac{\partial h^{-1}}{\partial z} \end{bmatrix} = \begin{bmatrix} \mathbf{o} \\ G_z \end{bmatrix} \quad (1.17)$$

where matrices  $G_x$  and  $G_z$  are the Jacobian of function  $h^{-1}$  with respect to the state vector and the observation vector respectively. By substituting the Jacobians in equations (1.16) and (1.17) into equation (1.15), the following matrix is obtained:

$$\hat{\Sigma} = \begin{bmatrix} \Sigma & \Sigma G_x^T \\ G_x \Sigma & G_x \Sigma G_x^T + G_z Q_t G_z^T \end{bmatrix} \quad (1.18)$$

The linearized covariance update can be factored as shown in equation (1.19) by assuming that  $A \equiv 1$ ,  $B \equiv 0$ ,  $C \equiv 0$  and  $D \equiv G_z$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^T = \begin{bmatrix} APA^T + BQB^T & APC^T + BQD^T \\ CPA^T + DQB^T & CPC^T + DQD^T \end{bmatrix} \quad (1.19)$$

$$\hat{\Sigma} = \begin{bmatrix} \mathbf{1} & \mathbf{o} \\ G_x & G_z \end{bmatrix} \begin{bmatrix} \Sigma & \mathbf{o} \\ \mathbf{o} & Q_t \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{o} \\ G_x & G_z \end{bmatrix}^T \quad (1.20)$$

$$(1.21)$$

## 1.5 NORMALIZED ESTIMATION ERROR SQUARED

The Mahalanobis distance is a measure of the distance between a point and a distribution. It is a way to measure how many standard deviations is away the point from the mean of the distribution. The Mahalanobis distance of an observation  $x = [x_1 \ x_2 \ \dots]$  with mean  $\mu = [\mu_1 \ \mu_2 \ \dots]$ , and covariance matrix  $\Sigma$  is defined as:

$$D = \sqrt{(x - \mu) \Sigma^{-1} (x - \mu)} \quad (1.22)$$

Every time the drone observes a pole or a marker, a node responsible of identifying the landmark processes the features that distinguish it with respect to other landmarks, and publishes these features as a message. In the case of poles, the message type is `RangeAndBearingPole`, which contains the range and bearing information, and in the case of markers the message type is `Pose`, which contains the position and orientation information of the observed marker. Given this, each observation  $z$  for a pole, is composed by three measurements, and each observation of a marker is composed by six measurements. After observing a landmark the Algorithm (1.3) calculates the discrepancy between the observation  $i$  and the predicted observation by the innovation ( $v^i$ ), and its covariance ( $S$ ), as

$$\begin{aligned} v^i &= z_t^i - h^i(\hat{\mu}_t) \\ S &= H_t^i \hat{\Sigma}_t H_t^{iT} + Q_t \end{aligned}$$

The square of the Mahalanobis distance can be used in order to establish a correspondence between the observed measurements with the landmark features if the following is hold:

$$D_i^2 = v^i S^{-1} v^i < \chi_{d,1-\alpha}^2 \quad (1.23)$$

where  $d$  is the size of  $h^i$  and  $1 - \alpha$  is the desired confidence level. This test is called individual compatibility, and when applied to the predicted state can be used to determine the subset of observation features that are compatible with the observation.

Furthermore, a state estimate is consistent if its state estimation error  $x - \hat{x}$  satisfies

$$\begin{aligned} \mathbb{E}[x - \hat{x}] &= 0 \\ \mathbb{E}[(x - \hat{x})(x - \hat{x})^T] &\leq \Sigma \end{aligned}$$

When the ground truth  $x$  is available, a Normalized Estimation Error Squared (NEES) can be performed to check the consistency of the filter. NEES test can be defined as the squared Mahalanobis distance for the difference between  $x$  and  $\hat{x}$ , and consistency can be checked with a  $\chi^2$  test

$$NEES = (x - \hat{x}) \Sigma^{-1} (x - \hat{x}) \leq \chi_{d,1-\alpha}^2$$

However, the ground truth is only available in simulated environments thus, the consistency of the filter is maintained by using observations that satisfy the test in equation (1.23). This way, the filter will discard observations that do not satisfy the innovation test, maintaining a consistent estimation of the state and therefore, the map.





---

## EKF-SLAM IMPLEMENTATION

---

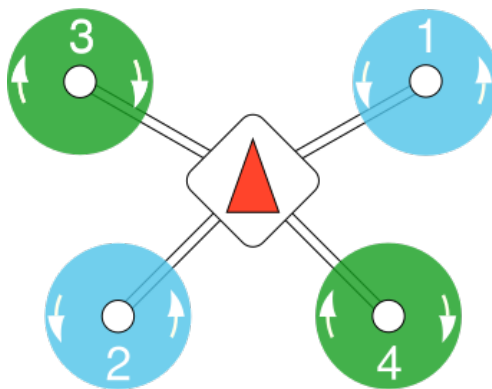
In Chapter 1 the EKF-SLAM algorithm in the context of SLAM was explained. As mentioned in Section 1.3, the algorithm can be summarized in two steps: prediction and correction. In the first stage involves the prediction of the next state of the system, meanwhile, during the correction stage, this estimation is updated.

In this chapter, an EKF-SLAM implementation is shown, starting from the used drone characteristics, going through the prediction and correction steps, and ending with the overall system's architecture.

### 2.1 THE DRONE

#### 2.1.1 Characteristics

The drone considered for this analysis has a common characteristics, such as four rotors disposed as an X. The drone frame is called 3DR Iris Quadrotor, and can be seen in Figure 2.1.



**Figure 2.1:** 3DR Iris Quadrotor frame.

#### 2.1.1.1 *Flight Controller*

The flight controller used in this case is a PixHawk 4, and its characteristics are shown in Table 2.1. It is worth mentioning that it includes an integrated accelerometer, a gyroscope, a magnetometer and a barometer.

#### 2.1.1.2 *Additional Sensors*

The drone was equipped with several sensors that are the input for localization, mapping and path planning algorithms, among others.

The localization and mapping algorithm makes use, in an indirect way, of monocular and stereo cameras, and laser range sensors. Four cameras were mounted in order to be able to have 360 range view, with cameras mounted every 90 degrees. Furthermore, two stereo cameras were mounted, one points forward in order to update the Octomap and the other points downwards in order to see the markers; Also, both of them are used to build a 3-dimensional map, used for obstacle avoidance and with the height estimation algorithm. Furthermore, eight laser range sensors were mounted every 45 degrees, and one PIX4Flow optical flow camera points downwards in order to measure the distance to the ground.

#### 2.1.2 *Reference Frames*

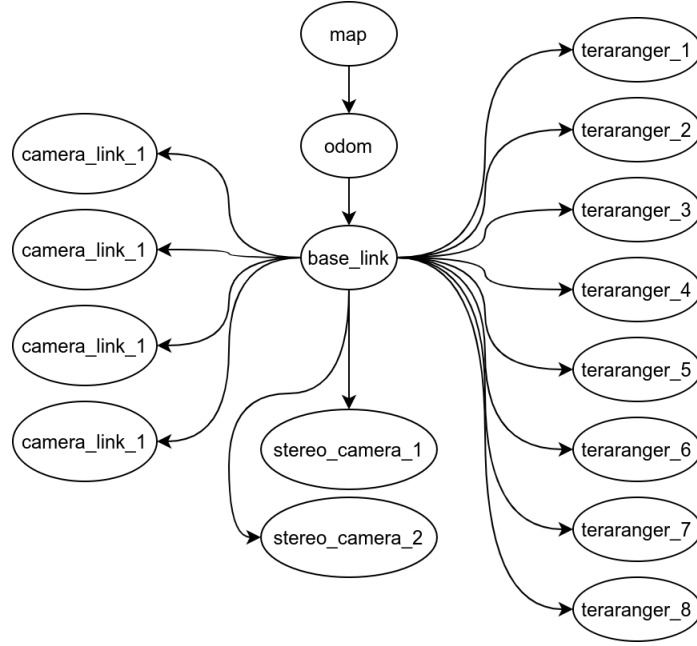
This system is composed by three reference frames linked with each other, as represented in Figure 2.2: *map*, *odom* and *base\_link*. The *map* frame, also called *world* or *global* frame, is the static reference frame, where the global drone's position and global markers' position is set. The *odom* reference frame is similar to the *map* frame, but the difference is that this frame drifts with time, as happens with the pure odometry. Finally, the *base\_link* frame, also referred as *body* or *local* frame, refers to the center of mass of the drone.

As mentioned before, all transformations within these reference frames are published by different nodes: the *map* to *odom* transformation is handled by the *rtabmap* node, the *odom* to *base\_link* is handled by *mavros* node. There are many other transformations in the system, but they are mainly related to the *base\_link* reference frame and the different cameras and sensors in the drone.

Finally, there is a transformation that worth mentioning: the *NED* to *ENU*. As mentioned in Section 1.1, ROS uses the *ENU* convention, while the convention adopted by the Aerospace community is the *NED*. Also, as mentioned in Section 1.1.5, MAVROS is in charge of doing and publish this transformation.

ITEM	DESCRIPTION
Main FMU processor	STM32F765 32 Bit Arm® Cortex®-M7, 216MHz, 2MB memory, 512KB RAM
IO Processor	STM32F100 32 Bit Arm® Cortex®-M3, 24MHz, 8KB SRAM
On-board sensors	Accel/Gyro: ICM-20689 Accel/Gyro: BMI055 Magnetometer: IST8310 Barometer: MS5611
GPS	ublox Neo-M8N GPS/GLONASS receiver; integrated magnetometer IST8310
Interfaces	8-16 PWM outputs (8 from IO, 8 from FMU) 3 dedicated PWM/Capture inputs on FMU Dedicated R/C input for CPM Dedicated R/C input for Spektrum / DSM and S.Bus with analog / PWM RSSI input Dedicated S.Bus servo output 5 general purpose serial ports 3 I2C ports 4 SPI buses Up to 2 CANBuses for dual CAN with serial ESC Analog inputs for voltage / current of 2 batteries
Power System	Power module output: 4.9 5.5V USB Power Input: 4.75 5.25V Servo Rail Input: 0 36V
Weight and Dimensions	Weight: 15.8g Dimensions: 44x84x12mm
Operating temperature	-40 to 85°C

**Table 2.1:** Technical specification of the PixHawk 4 flight controller.



**Figure 2.2:** Main reference frames in the system.

In Figure 2.3 a simple scheme of the transformation can be seen. As explained before, this transform consists on rotating the X-axis and the Y-axis by 90 degrees, hence the homogeneous rotation matrix will have the following form:

$$R_{ENU}^{NED} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.1)$$

## 2.2 MOTION MODEL

During the prediction step, the motion model and the covariance matrix update are computed. The motion model will update the state vector, which will store the position X, Y and Z in the global reference frame, and the orientation in the Z-axis of the drone. Fortunately MAVROS provides a node that makes odometry estimation based on different sensors outputs. The messages published by the odometry node, of type `Odometry`, provide the linear velocity, the angular velocity and pose information. The velocity information is used to estimate the position in X, Y and Z, and the drone's orientation along the Z-axis. However, the velocity estimation provided by MAVROS is relative to the body reference frame, and this has to be transformed into the world

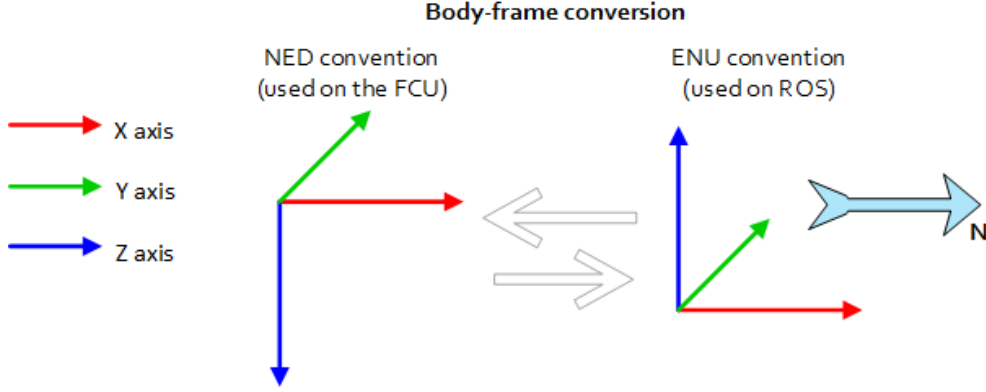


Figure 2.3: NED to ENU conversion scheme. [9]

reference frame, so before estimating the global position of the drone it is mandatory to do this transformation.

$$\mathbf{u} = \begin{bmatrix} v_x^b & v_y^b & v_z^b & \omega_x^b & \omega_y^b & \omega_z^b & \phi^b & \theta^b & \psi^b \end{bmatrix}^T \quad (2.2)$$

$$\mathbf{v}^w = \mathbf{T} * \begin{bmatrix} v_x^b \\ v_y^b \\ v_z^b \end{bmatrix} \quad (2.3)$$

where the control vector  $\mathbf{u}$  is composed by the linear velocities in the body reference frame  $(v_x^b, v_y^b, v_z^b)$ , the angular velocities in the body reference frame  $(\omega_x^b, \omega_y^b, \omega_z^b)$  and the drone's orientation  $(\phi^b, \theta^b, \psi^b)$ . Additionally,  $\mathbf{T}$  is the homogeneous transform (see Section 1.2) using the current orientation of the drone:  $\phi^b, \theta^b$  and  $\mu_\psi$ . Notice the third element of the orientation ( $\mu_\psi$ ), this is because the orientation over the Z-axis is estimated by the filter. As result,  $\mathbf{v}^w$  will be a column vector with the linear velocities in the global reference frame.

Given this, the motion model update can be summarized in the following calculation:

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} \mu_{t-1,x^w} \\ \mu_{t-1,y^w} \\ \mu_{t-1,z^w} \end{bmatrix} + \Delta t * \mathbf{v}^w \quad (2.4)$$

$$\hat{\mu}_\psi = \mu_{t-1,\psi} + \Delta t * \omega_z^b \quad (2.5)$$

Then,  $G_t$ , which is the Jacobian matrix of the motion model, should be computed. As mentioned in Section 1.4.1, it has the following characteristic:

$$G_t = \begin{bmatrix} G_r & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (2.6)$$

$$G_r = \begin{bmatrix} 1 & 0 & 0 & G_{r,14} \\ 0 & 1 & 0 & G_{r,24} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.7)$$

$$\begin{aligned} G_{r,14} &= -\Delta t * (v_y^w * (c_\phi * c_\psi + s_\theta * s_\phi * s_\psi) - v_z^w * (c_\psi * s_\phi - c_\phi * s_\theta * s_\psi) + v_x^w * c_\theta * s_\psi) \\ G_{r,24} &= -\Delta t * (v_z^w * (c_\psi * s_\phi + c_\phi * s_\theta * c_\psi) - v_y^w * (c_\phi * s_\psi - s_\theta * s_\phi * c_\psi) + v_x^w * c_\theta * s_\psi) \end{aligned}$$

where

$$\begin{aligned} c_\phi &= \cos(u_\phi), & c_\theta &= \cos(u_\theta), & c_\psi &= \cos(\mu_\psi) \\ s_\phi &= \sin(u_\phi), & s_\theta &= \sin(u_\theta), & s_\psi &= \sin(\mu_\psi) \end{aligned}$$

The function  $g$  that models the motion of the drone is assumed to be perfect and therefore noise free. So, before computing the covariance update, it is necessary to compute the process noise covariance matrix ( $R_t$ ) that encodes the motion model noise which, in this case, is related to the underlying dynamics of the drone flight. The noise is assumed to be additive and Gaussian, and therefore, the motion model can be decomposed as:

$$x_t = g(u_t, x_{t-1}) + \mathcal{N}(0, R_t) \quad (2.8)$$

$$R_t = N * U * N^T \quad (2.9)$$

the noise part in equation (2.8) relates to the acceleration component that is, in this case, unknown. However, it is known from a theoretical perspective: the acceleration component in an accelerated movement is  $\frac{1}{2}\Delta t^2 a$ , where  $a$  is the body's acceleration. Given this, we can assume that the noise component in the motion model is:

$$\mathcal{N}(0, R_t) = \frac{1}{2}\Delta t^2 T_w^b \begin{bmatrix} a_x \\ a_y \\ a_z \\ a_\psi \end{bmatrix} \quad (2.10)$$

where  $T_w^b$  is the transformation matrix from body to world reference frame. The covariance of the process noise ( $R_t$ ) can be decompose as shown in

equation (2.9), where matrix  $N$  is the Jacobian of acceleration term with respect to the state vector, and matrix  $U$  is the estimated average acceleration.

$$N = \frac{\partial \frac{1}{2} \Delta t^2 T_w^b a}{\partial \mu} \quad (2.11)$$

$$U = I * \begin{bmatrix} a_{avg,x} \\ a_{avg,y} \\ a_{avg,z} \\ a_{avg,\psi} \end{bmatrix} \quad (2.12)$$

The multiplication in (2.9) provides an approximate mapping between the motion noise in control space and the motion noise in the state space.

Finally, the covariance update should be computed as follow

$$\hat{\Sigma} = G_t * \Sigma * G_t^T + R_t \quad (2.13)$$

### 2.3 OBSERVATION MODELS

While the drone is moving around the environment it senses different landmarks that may be included or not in the state vector. These observations will eventually improve the localization of the drone, and will improve the landmarks' poses if needed. The whole process involves the computation of the Jacobian of the observation model for the seen landmark, the computation of the Kalman gain, and the update of the state vector and covariance matrix.

To perform the correction step, EKF-SLAM needs a linearized observation model with additive Gaussian noise. In the case studied in this work, there are two kinds of landmarks and, therefore, two different observation models. The main difference between these two type of landmarks, is that the position of poles type is known, while it is not known in the case of markers. Hence, every time the robot "sees" a pole, the algorithm will update the drone's pose and the known markers' pose; while every time it "sees" a marker two course of action are possible:

- a) if the marker is not known, it is added to the state vector, enlarging it along with the covariance matrix.
- b) if the marker is known, its pose, the pose of all other markers, and the drone's pose are updated.

The observation model is, as with the motion model in equation (2.8), assumed to be perfect and with an additive Gaussian noise. The noise here is related to the observation process, and so, related to the used sensors.

$$z_i = h_i(x_t) + \mathcal{N}(0, Q_t) \quad (2.14)$$

Consequently, the noise covariance matrix of the observation model cannot be deducted as with the noise covariance matrix of the motion model. In this case, the matrix should be constructed empirically based on the sensors' characteristics, and it has the following characteristic:

$$Q_t = \begin{bmatrix} \sigma_1^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma_n^2 \end{bmatrix} \quad (2.15)$$

where the diagonal elements  $\sigma_{1..n}$  are the standard deviation of the sensor. Depending on the sensor used, the diagonal elements can be the standard deviation for the range and bearing components (distance, azimuth and elevation), or others.

As shown in Algorithm 1.3 several steps are followed during the correction part of the algorithm. After computing the observation model and its Jacobian matrix, the Kalman gain and the innovation should be calculated, and finally, the state vector and covariance matrix updates should be done.

### 2.3.1 Observation model for laser range

Observation model of laser range

### 2.3.2 Observation model for Poles

In the case of Poles, a range and bearing method is used. In this case, since the poles have a known position, their information is not kept in the state vector and therefore, this information will be used for localization purposes.

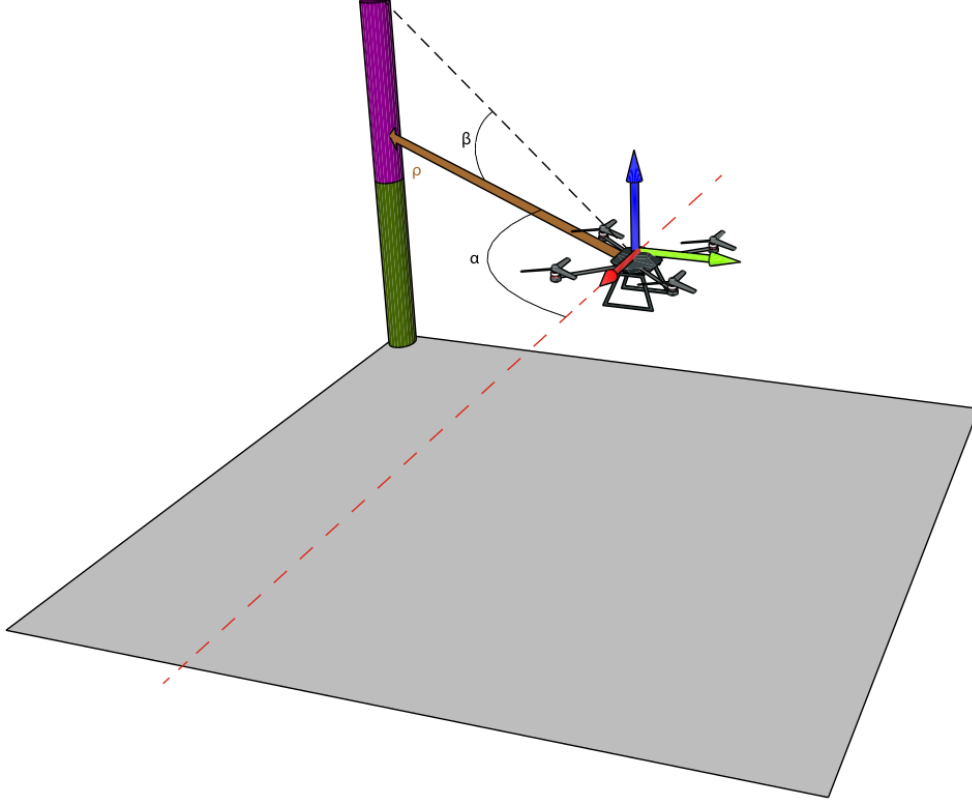
A ROS node will publish the range and bearing information every time the drone sees a pole, and this information will be used to calculate the innovation based on the predicted range and bearing. Hence, the observation model used for poles is computed in the following way:

$$\begin{bmatrix} p_{i,x}^b \\ p_{i,y}^b \\ p_{i,z}^b \end{bmatrix} = T_r^{-1} \begin{bmatrix} p_{i,x}^w \\ p_{i,y}^w \\ p_{i,z}^w \end{bmatrix} \quad (2.16)$$

$$h_i(\hat{\mu}_t) = \begin{bmatrix} p_{i,\rho} \\ p_{i,\alpha} \\ p_{i,\beta} \end{bmatrix} = \begin{bmatrix} \sqrt{p_{j,x^b}^2 + p_{j,y^b}^2} \\ \text{atan2}(p_{j,y}^b, p_{j,x}^b) \\ \text{atan2}(p_{j,z}^b, p_{j,\rho}^b) \end{bmatrix} \quad (2.17)$$



In equation (2.16),  $T_r^{-1}$  corresponds to the inverse of the homogeneous transformation matrix with respect to the current drone's pose, and elements  $p_i^w$  are the  $x$ ,  $y$  and  $z$  coordinates of the  $i$  pole's tip in the world reference frame. This way, the global position of the pole  $i$  is projected to the body reference frame. After that, the range ( $\rho$ ), azimuth angle ( $\alpha$ ) and elevation angle ( $\beta$ ) are calculated, as shown in equation (2.17). In Figure 2.4 an example of the range and bearing is shown.

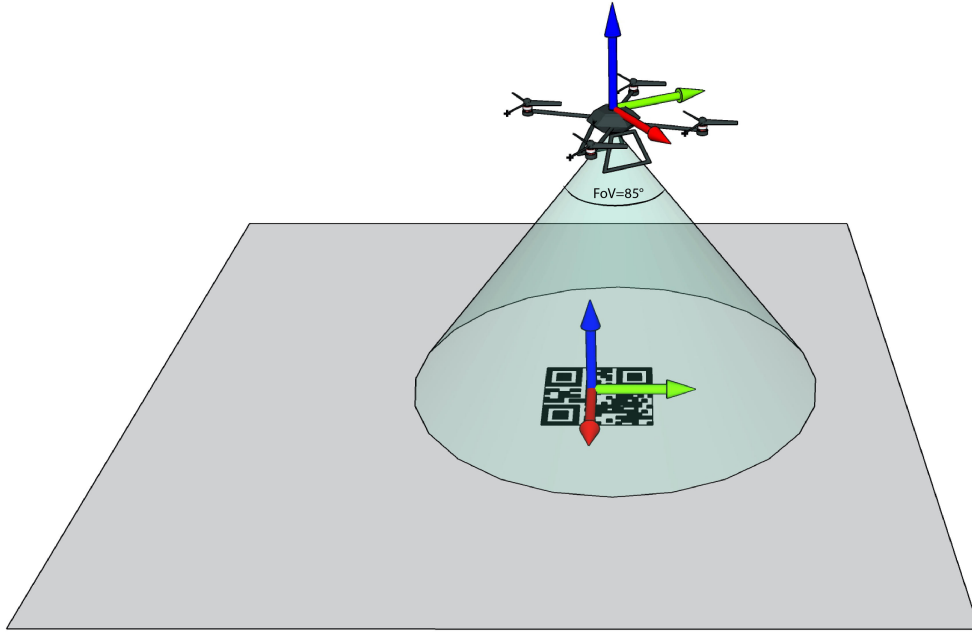


**Figure 2.4:** Range and bearing example. The drone observes a pole, process the data from the sensors and estimates the distance ( $\rho$ ), the elevation angle ( $\beta$ ), and the azimuth angle ( $\alpha$ ). The elevation angle is calculated based on the top extreme of the pole.

Observing a pole affects only the drone's pose, and therefore, the Jacobian matrix of the observation model will have the following form:

$$H_i = \begin{bmatrix} \frac{\partial \rho'}{\partial \mu_x} & \frac{\partial \rho'}{\partial \mu_y} & \frac{\partial \rho'}{\partial \mu_z} & \frac{\partial \rho'}{\partial \mu_\psi} & \dots & 0 \\ \frac{\partial \alpha'}{\partial \mu_x} & \frac{\partial \alpha'}{\partial \mu_y} & \frac{\partial \alpha'}{\partial \mu_z} & \frac{\partial \alpha'}{\partial \mu_\psi} & \dots & 0 \\ \frac{\partial \beta'}{\partial \mu_x} & \frac{\partial \beta'}{\partial \mu_y} & \frac{\partial \beta'}{\partial \mu_z} & \frac{\partial \beta'}{\partial \mu_\psi} & \dots & 0 \end{bmatrix} \quad (2.18)$$

where  $\rho'$  corresponds to the distance part of the observation model,  $\alpha'$  is the azimuth part, and  $\beta'$  the elevation part. The elements after the 4<sup>th</sup> column



**Figure 2.5:** Example of the drone observing a marker. The camera visualizes a marker and the node `visp_auto_tracker` estimates its pose with respect to the camera reference frame.

are all 0, which means, as said before, that the observation of a pole will not affect the pose of the markers.

### 2.3.3 Observation model for Markers

The observation model for the markers is a bit different. In this case a ROS node is responsible of detecting, tracking and publishing the pose of the markers with respect to the camera that has seen it. The ROS package responsible of this process is called `visp_auto_tracker`, and publishes messages of type `PoseStamped`, which provides the position and orientation of the seen marker. An example of this situation can be seen in Figure 2.5.

Every time the camera that points down sees a known marker, the `visp_auto_tracker` node will publish the pose of that marker. This published pose is with respect to the camera which is not positioned at the drone's center of mass and so, a different transformation is needed. This process can be seen in equation (2.19),

in which the transformation of the marker's pose in the world reference frame is transformed to the pose in the camera reference frame.

$$\begin{bmatrix} m_{i,x}^c \\ m_{i,y}^c \\ m_{i,z}^c \\ m_{i,\phi}^c \\ m_{i,\theta}^c \\ m_{i,\psi}^c \end{bmatrix} = (T_r * T_c)^{-1} * T_m \quad (2.19)$$

where,  $T_r$  is the homogeneous transformation matrix of the drone's pose,  $T_c$  is the homogeneous transformation matrix of the camera's pose, and  $T_m$  is the homogeneous transformation of the marker's pose. The result is a vector that contains the marker's pose in the camera reference frame.

As mentioned before, observing a marker will update the drone's pose and that marker's pose, and therefore the Jacobian matrix of the observation model will have a different aspect from the pole's case. Here, the Jacobian can be split in two parts as shown in equation (2.20), where the left part, as with the poles, affects the drone's pose, while the right part of the matrix affects the seen marker's pose.

$$H_i = \begin{bmatrix} \frac{\partial h_i(\hat{\mu})}{\partial \mu_r} & \dots 0 \dots & \frac{\partial h_i(\hat{\mu})}{\partial \mu_{m_i}} & \dots 0 \dots \end{bmatrix} \quad (2.20)$$

The Jacobian matrices of the observation model with respect to the drone state and with respect the marker pose can be seen in Appendix A. It is worth to mention that the reason why this matrix only contains two sections different to zero is because of the relations between markers: observing a marker does not affect the pose of others, and therefore, the Jacobians of the observation model for marker  $i$  with respect to marker  $i + 1$  or any other marker is 0.

Furthermore, unlike the case of poles, the markers' pose is unknown the first time, and therefore the algorithm will introduce their pose into the state vector when the drone sees a previously unknown marker.

#### 2.3.3.1 Adding new Markers

As mentioned in Section 1.4.2, to add new landmarks to the state vector an inverse observation model is needed. In the current case, the inverse observation model will project the observed marker's pose from the camera reference frame to the world reference frame. Equation 2.21 shows the inverse observation model for markers where, similarly to the observation model, a

set of transformations are performed but, this time, in order to obtain the pose of the seen marker in the world reference frame:

$$\begin{bmatrix} m_{i,x}^w \\ m_{i,y}^w \\ m_{i,z}^w \\ m_{i,\phi}^w \\ m_{i,\theta}^w \\ m_{i,\psi}^w \end{bmatrix} = T_r * T_c * T_m \quad (2.21)$$

As before,  $T_r$  is the homogeneous transformation matrix of the drone's pose,  $T_c$  is the homogeneous transformation matrix of the camera's pose, and  $T_m$  is the homogeneous transformation of the marker's pose. The result is a vector containing the marker's pose in the world reference frame, and with which will extend the state vector:

$$\mu_t = \left[ \mu_t \mid m_{i,x}^w \quad m_{i,y}^w \quad m_{i,z}^w \quad m_{i,\phi}^w \quad m_{i,\theta}^w \quad m_{i,\psi}^w \right]^T$$

Furthermore, the covariance matrix should be updated in order to contain the newly added marker. As shown in equation 1.15, the Jacobian of the inverse observation model with respect to the drone's state and the Jacobian of the inverse observation model with respect to the marker's state are needed. Both Jacobian matrices can be seen in Appendix A. In this case, the matrix  $Y_x$  has a size of  $(|\mu| + 6) \times (|\mu|)$  where  $|\mu|$  is the size of the state vector before adding the projected pose of the seen marker. On the other hand, the matrix  $Y_z$  has a size of  $(|\mu| + 6) \times 6$ . Given these two matrices, the new covariance matrix  $\Sigma$  will be increased by 6 columns and 6 rows and will have the following form:

$$\Sigma = \begin{bmatrix} \Sigma_{r,r} & \Sigma_{r,m} \\ \Sigma_{r,m} & \Sigma_{m,m} \end{bmatrix}$$

where  $\Sigma_{r,r}$  is the covariance matrix between the drone's state variables, the  $\Sigma_{r,m}$  and  $\Sigma_{m,r}$  is the covariance matrix between the drone and the markers' state variables, and between the markers and the drone's state variables, and  $\Sigma_{m,m}$  is the covariance matrix between markers' state variables.

## 2.4 VISUAL PROCESS PIPELINE

As mentioned in Section 2.1, the drone is composed by several cameras, specifically, two stereo cameras and four monocular cameras. These cameras are used to identify and avoid obstacles and for localization purposes. The localization process involves the identification of poles and markers, and

therefore, the usage of the before mentioned cameras: the four monocular cameras are used to identify poles, while one stereo camera is used to identify the markers.

#### 2.4.1 Visual process for Poles

Four ROS nodes are responsible of determining whether a pole is being seen or not. To do so, they subscribe to topics related to the camera information and the camera image, and every time a new image arrives, each node process it to identify a pole. Each pole is composed by two different colors, and each pole has a different combination of colors making possible to uniquely identify them. The overall process can be followed in Algorithm (2.1).

---

#### Algorithm 2.1: Poles localization algorithm

---

**Input:** image\_raw, camera\_info

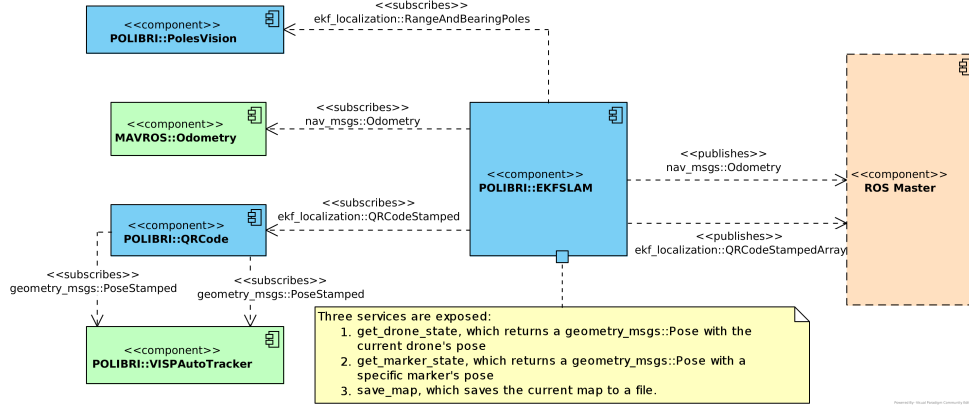
```

1 Apply masks to image
2 Identify how many poles are in the image
3 if contours of poles where found then
4   foreach pole contour do
5     Find coordinates of contour
6     Identify pole
7     if the full width of the pole can be seen then
8       Compute Azimuth ( $\alpha$ )
9       if the pole is not occluded then
10        Compute Range ( $\rho$ )
11        if the top of the pole can be seen then
12          Compute Altitude ( $\beta$ )
13 return  $\alpha$ ,  $\beta$ ,  $\rho$ 

```

---

Lines 1 and 2, applies a mask in order to identify the amount of poles contours in the image. If there are pole contours in the image, at lines 5 and 6 the algorithm uniquely identifies the pole and its coordinates. If the width of the identified pole can be seen, at line 8 the algorithm computes the azimuth angle ( $\alpha$ ) by using the drone's yaw and the center of the contour. At line 9, it estimates the probability of occlusion of the pole by comparing the contour shape with the form factor, computed as  $\frac{y_{\max} - y_{\min}}{x_{\max} - x_{\min}}$  where  $y_{\max}$  and  $y_{\min}$  are the bottom and the top coordinates of the contour, and  $x_{\max}$  and  $x_{\min}$  are the right and left coordinates of the contour. If the probability is greater that 75%, the range is computed based on the diameter of the contour. Finally,



**Figure 2.6:** Components diagram of the EKF Localization node. The EKFSLAM class subscribes to different messages, between them the Odometry messages are used as control variables, while the RangeAndBearingPole and QRCodeStamped messages are used every time the drone observes a pole or a marker. Moreover, EKFSLAM class publishes two types of messages: Odometry which provides the filtered localization and QRCodeStampedArray which contains a list markers' poses.

if the top of the contour can be seen (line 11), at line 12 it computes the altitude.

#### 2.4.2 Visual process for Markers

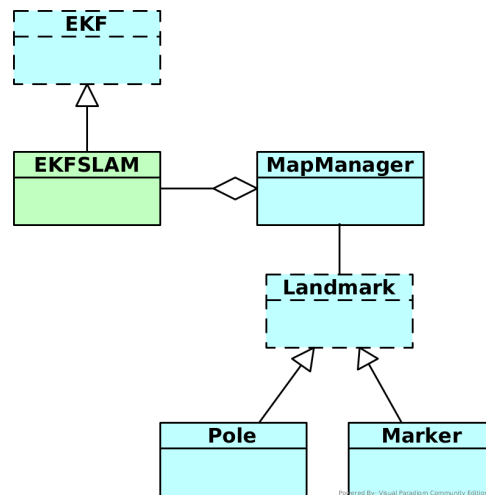
The markers detection pipeline is performed by the Visual Servoing Platform (ViSP) tracking package [17]. The package wraps an automated pattern-based tracker based on ViSP library, which allows to estimate an object pose with respect to the camera. In the case of this work a QR code is used as marker, which is identified by the tracker.

The `visp_auto_tracker` node subscribes to the camera image and information topics, and publishes the detected object's pose. The camera messages are provided by the stereo camera that points downwards.

### 2.5 OVERALL ARCHITECTURE

The system is composed by several ROS nodes that specialize and perform a wide range of operations. However, what is most interesting for this work is the node that is in charge of the EKF-SLAM process.

In Figure 2.6 the components that interact around the EKF Localization node can be seen. In the center of the figure the EKFSLAM component is placed, which is responsible of running the EKF-SLAM algorithm. Every



**Figure 2.7:** Class diagram of the EKFSLAM node. The EKFSLAM class is composed by a MapManager instance, which is composed by a list of Landmarks. The Landmark class is abstract and its concrete classes are Pole and Marker.

time an Odometry message arrives, the prediction step takes place, and this happens every 30Hz. Furthermore, every time a RangeAndBearingPoles or a QRCodeStamped arrives the correction step takes place, and differently to the prediction step, this depends on whether the drone observes or not a pole or a marker. Additionally, the EKFSLAM component provides three services:

1. `get_drone_state`: takes no arguments, and returns the current localization of the drone Pose message.
2. `get_marker_state`: takes as argument the id of a marker, and returns the current estimated pose of that markerPose message.
3. `save_map`: takes no arguments, and it saves the current map in a YAML file. The map contains the pose of all the poles and all the markers seen so far.

It is worth to mention that the messages of type QRCodeStamped are provided by the QRCode node which is internal to the `ekf_localization` package. This node is needed because the `visp_auto_tracker`, responsible of identifying the markers and publish their poses, publish these data as separate messages: the marker's id as a String message and the pose as a PoseStamped message. Due to this situation, the QRCode node is responsible of match both data together and publish an QRCodeStamped message which contains the marker's identifier and its pose.

The main element of the EKF-SLAM node is the EKFSLAM class which is responsible of keeping track of the state vector and execute the EKF-SLAM

algorithm. This class inherits from an abstract class called EKF, which is responsible of the implementation of the algorithm in a generic way, while the specifics for the current problem is contained in its child.

There are several components that interact within EKFSLAM class, but probably the most interesting one is the MapManager class. This class is responsible of maintaining an updated map of the environment with all the poles and markers, and to save and retrieve the map from a YAML file.

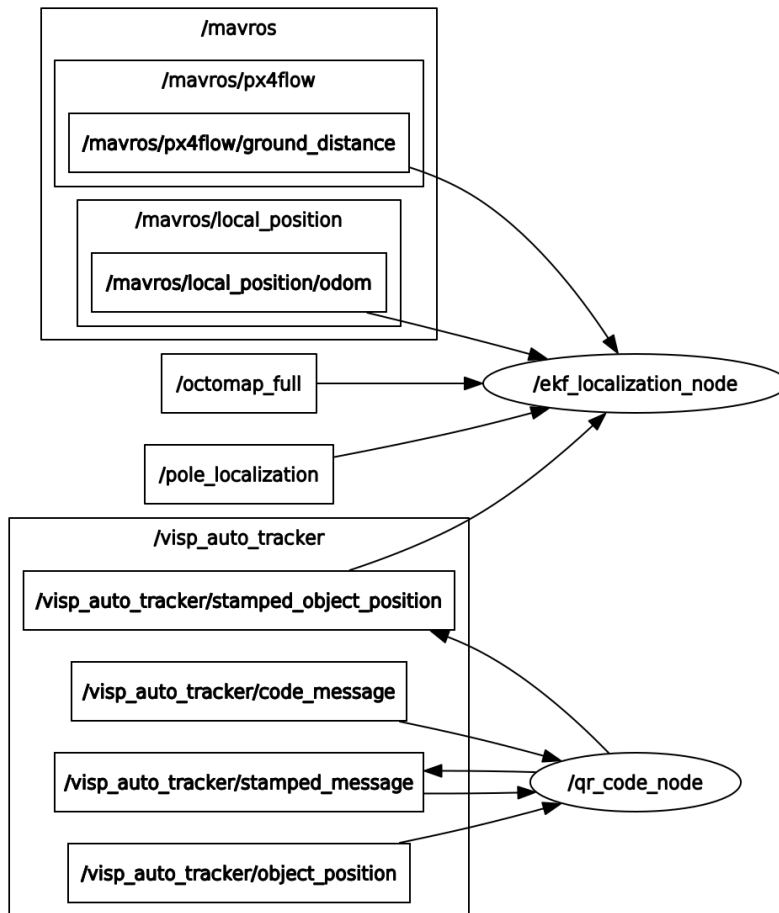
The MapManager class contains a list of Landmarks, and every time the state vector is updated in EKFSLAM class, the MapManager class updates the information about each Marker. As mentioned before, the pose of each pole is known and needs no update. Furthermore, the Landmark class is specialized in a Marker class and a Pole class, each of which is responsible of provide the observation model associated to it and the needed Jacobian matrices for the algorithm, hiding the implementation to the EKFSLAM class.

### 2.5.1 ROS nodes

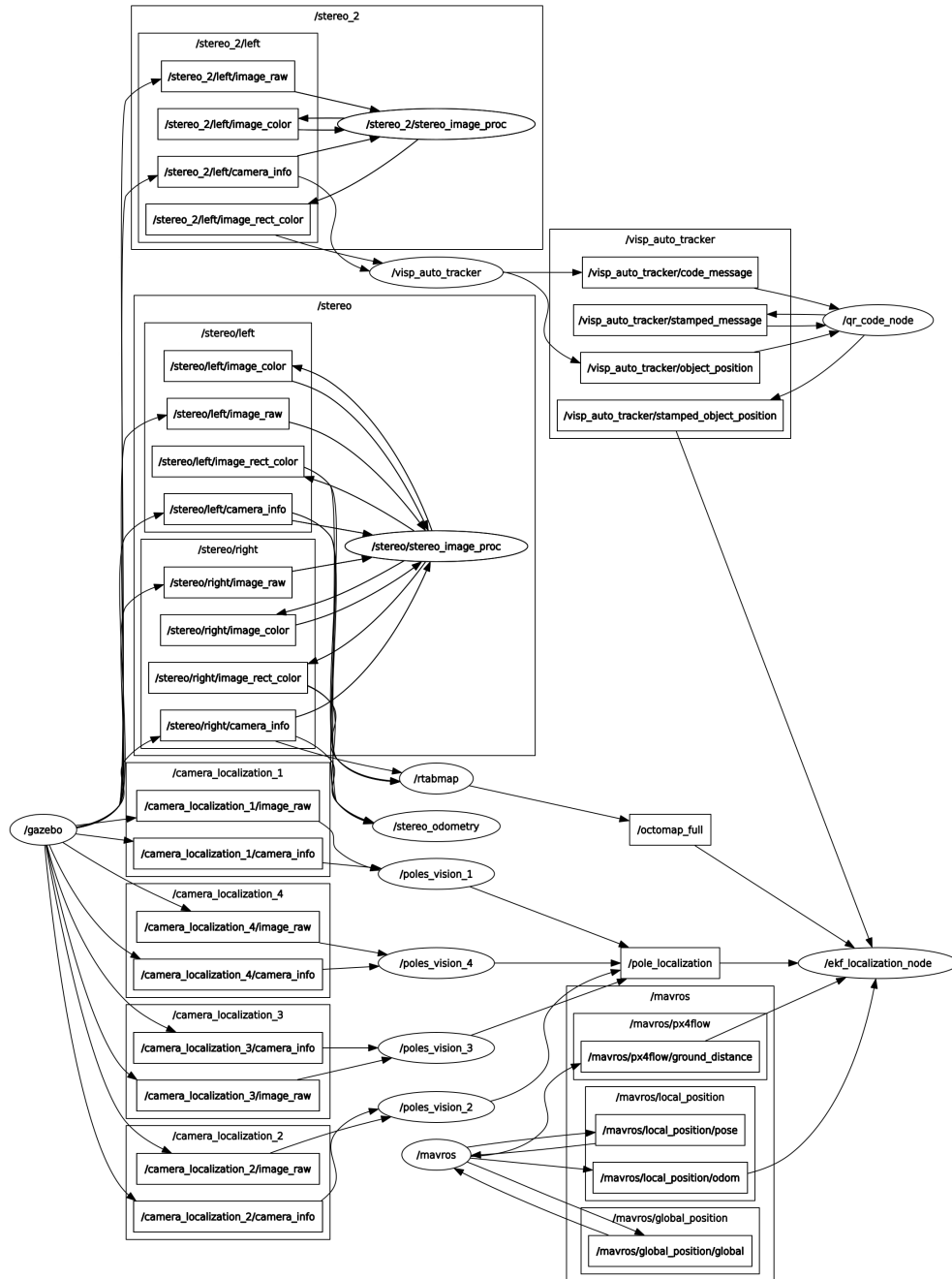
Figure 2.8 shows the detail for the localization node, /ekf\_localization\_node, with all its subscriptions. It is worth to mention the /qr\_code\_node, which publishes the markers position with its identifier, subscribes to /visp\_auto\_tracker/code\_message and /visp\_auto\_tracker/object\_position, and publishes into /visp\_auto\_tracker/stamped\_object\_position topic.

Regarding the ROS nodes that comprise the system, in Figure 2.9 all the ones that interact with the localization node and all the others are shown. The EKF-SLAM node, and as mentioned before, interacts with /mavros node, all the poles identification nodes (/poles\_vision\_#), the /qr\_code\_node, and the /rtabmap node that provides the Octomap messages. The message passing is represented by arrows, where an incoming arrow means that a node subscribes to the message topic, while an outgoing arrow means that the node publishes that message topic. Arrows go from a node to a topic or from a topic to a node, but do not go from node to node. Its reason lies on the fact that nodes interact between each other using a message passing interface as mention in Section 1.1. Furthermore, it is worth to mention that in Figure 2.9 a /gazebo node is present. This node appears only on simulation and represents the simulated environment, that is why it publishes all the cameras and stereo cameras information.





**Figure 2.8:** Detail of the EKF-SLAM node interactions. Beside the `/ekf_localization_node`, the `/qr_node_code` can be seen. This node, is responsible of subscribing the `/visp_auto_tracker` messages in order to, then, publish the identifier along with the pose of the marker that has been seen. On the other hand, the `/ekf_localization_node` subscribes to the `/mavros`, `/octomap`, `/pole_localization` and `/visp_auto_tracker` messages.



**Figure 2.9:** ROS graph of the system. Nodes are depicted as ellipses, while message topics are depicted as boxes. Each arrow means that a node publishes or subscribes to a specific message topic.

---

## EXPERIMENTAL RESULTS

---

Introduction where I should explain the purpose of the experiments.

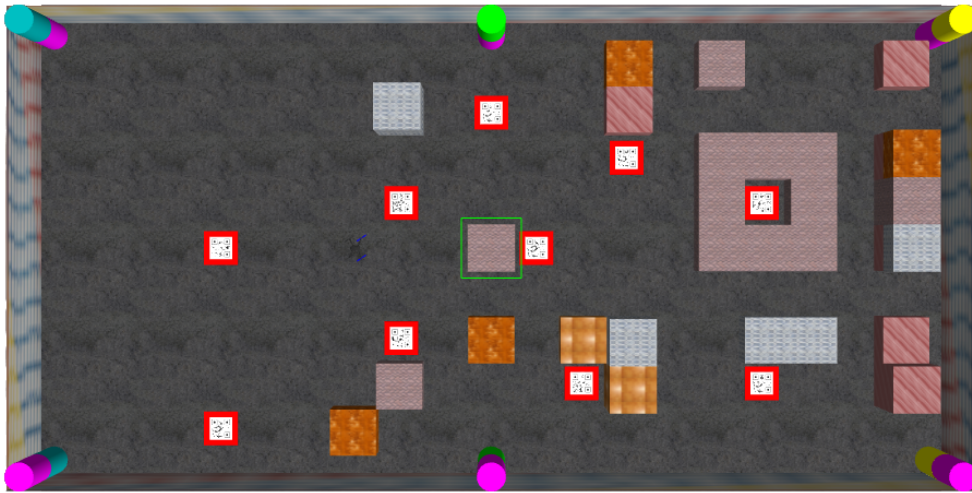
### 3.1 THE ENVIRONMENT

The simulated environment can be seen in Figure 3.1, where all the obstacles, poles and markers are disposed. Moreover, the walls define the limits of the environment, making the drone impossible to go off the limits. The markers were disposed arbitrarily around the world, while the poles are in the same place as they should be in the competition: one in each corner, and two along the middle axis of the space. The drone is placed at coordinates  $(-3;0)$ , and can be depicted in the figure as the green circle in the middle right. As mentioned before, the poles have a different combination of colors, where two poles have not the same combination. The markers, are QR codes framed in red, while obstacles are brown and gray boxes around the environment. The walls are made of a net-like material distinguishable from the obstacles, while the floor is a pavement-like material.

### 3.2 SIMULATED EXPERIMENTS

#### 3.2.1 *Procedures*

#### 3.2.2 *Results*



**Figure 3.1:** Gazebo simulated environment.

---

## CONCLUSIONS AND FUTURE WORK

---

### 4.1 CONCLUSION

### 4.2 FUTURE WORK



---

## BIBLIOGRAPHY

---

- [1] *Robot Operating System*. Oct. 6, 2020. URL: <https://www.ros.org/> (visited on 10/06/2020) (cit. on p. 2).
- [2] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. Intelligent robotics and autonomous agents. The MIT Press, 2006. ISBN: 978-0-262-20162-9 (cit. on pp. 11, 12).
- [3] Howie Choset et al. *Principles of robot motion: theory, algorithms, and implementation*. Ed. by Ronald C. Arkin. Intelligent Robotics and Autonomous Agents. The MIT Press, 2005. ISBN: 9780262033275.
- [4] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems." In: *Journal of Basic Engineering* 82.1 (Mar. 1960), pp. 35–45. ISSN: 0021-9223. eprint: [https://asmedigitalcollection.asme.org/fluidengineering/article-pdf/82/1/35/5518977/35\\_1.pdf](https://asmedigitalcollection.asme.org/fluidengineering/article-pdf/82/1/35/5518977/35_1.pdf) (cit. on p. 8).
- [5] Roland Siegwart, Illah R. Nourbakhsh, and Davide Scaramuzza. *Introduction to autonomous mobile robots*. Ed. by Ronald C. Arkin. 2nd. Intelligent robotics and autonomous agents. The MIT Press, 2004. ISBN: 978-0-262-01535-6 (cit. on p. 14).
- [6] José A. Castellanos, José Neira, and Juan D. Tardós. *Autonomous Mobile Robots: Sensing, Control, Decision Making and Applications*. Ed. by FRANK L. LEWIS. CRC Press, 2006. Chap. 9. ISBN: 978-0-8493-3748-2.
- [7] Nikolaus Correl. *Introduction to Autonomous Robots*. Magellan Scientific, 2020. Chap. 11. ISBN: 978-0692700877.
- [8] *MAVLink*. Oct. 6, 2020. URL: <https://mavlink.io/en/> (visited on 10/06/2020).
- [9] *MAVROS*. Oct. 6, 2020. URL: <https://github.com/mavlink/mavros> (visited on 10/06/2020) (cit. on p. 25).
- [10] J.L. Blanco. *Derivation and implementation of a full 6D EKF-based solution to bearing-range SLAM*. Tech. rep. University of Malaga, 2008.
- [11] Peter Corke. *Robotics, Vision and Control*. Ed. by Bruno Siciliano and Oussama Khatib. Vol. 118. Springer, 2017. ISBN: 978-3-319-54412-0.
- [12] Basilio Bona. *Dynamic Modelling of mechatronic systems*. CELID, 2018. ISBN: 8867890115 (cit. on p. 9).

- [13] Mathieu Labbé and Francois Michaud. “RTAB-Map as an Open-Source Lidar and Visual SLAM Library for Large-Scale and Long-Term Online Operation.” In: *Journal of Field Robotics* (2018), pp. 1–31.
- [14] Armin Hornung, Kai M. Wurm, Maren Bennewitz, Cyrill Stachniss, and Wolfram Burgard. “OctoMap: An Efficient Probabilistic 3D Mapping Framework Based on Octrees.” In: *Autonomous Robots* (2013). Software available at <http://octomap.github.com> (cit. on p. 4).
- [15] Mathieu Labbé. “Simultaneous Localization and Mapping (SLAM) with RTAB-Map.” Québec, Nov. 2015 (cit. on p. 5).
- [16] Kris Hauser. *Robotic Systems*. 2018. URL: <http://motion.pratt.duke.edu/RoboticSystems>; <https://github.com/krishhauser/RoboticSystemsBook> (cit. on p. 8).
- [17] E. Marchand, F. Spindler, and F. Chaumette. “ViSP for visual servoing: a generic software platform with a wide class of robot control skills.” In: *IEEE Robotics and Automation Magazine* 12.4 (2005), pp. 40–52 (cit. on p. 34).





---

## DOCUMENTATION

---



B

---

USER MANUAL

---