Earth-fixed Trajectory and Map online estimation: building on GES sensor-based SLAM filters

Pedro Lourenço[a,b,*], Bruno J. Guerreiro[e,b], Pedro Batista[b,c], Paulo Oliveira[b,d], Carlos Silvestre[f,b]

[a] *GMV, Av. D. João II, Nº 43, T. Fernão de Magalhães, 7º, 1998-025 Lisboa, Portugal*
[b] *Institute for Systems and Robotics, Laboratory for Robotics and Engineering Systems, Portugal.*
[c] *Instituto Superior Técnico, Univ. de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*
[d] *Dep. of Mechanical Eng., Instituto Superior Técnico, Univ. de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*
[e] *Dep. of Electrical and Computer Eng., Univ. Nova de Lisboa, Caparica 2829-516, Portugal.*
[f] *Dep. of Electrical and Computer Eng., Faculty of Science and Technology, Univ. of Macau*

Abstract

This paper addresses the problem of obtaining an Earth-fixed trajectory and map (ETM), with the associated uncertainty, using the sensor-based map provided by a globally asymptotically/exponentially stable (GES) SLAM filter. The algorithm builds on an optimization problem with a closed-form solution, and its uncertainty description is derived resorting to perturbation theory. The combination of the algorithm proposed in this paper with sensor-based SLAM filtering results in a complete SLAM methodology, which is directly applied to the three main different formulations: range-and-bearing, range-only, and bearing-only. Simulation and experimental results for all these formulations are included in this work to illustrate the performance of the proposed algorithm under realistic conditions. The ETM algorithm proposed in this paper is truly sensor-agnostic, as it only requires a sensor-based map and imposes no constraints on how this map is acquired nor how egomotion is captured. However, in the experiments presented herein, all the sensor-based filters use a sensor to measure the angular velocity and, for the range-only and bearing-only formulations, a sensor to measure the linear velocity.

Keywords: SLAM, Procrustes Problem, Perturbation Theory, Mapping, Robotics.

1. Introduction

Search and rescue, surveillance, and the automatic inspection of critical infrastructures and buildings, such as bridges, electric power lines, dams, and construction sites, have been recently acknowledged as challenging and promising application scenarios for the use of unmanned aerial vehicles (UAVs) (see [1, 2, 3, 4, 5, 6] and references therein). Navigation and positioning systems are of the utmost importance in the development of these vehicles, particularly in mission scenarios where geo-referencing is not possible. Near these structures, the global positioning system (GPS) signal can be severely degraded and the magnetic field is locally distorted, precluding the use of GPS receivers and magnetometers. Therefore, either indoors or outdoors, relative positioning systems are fundamental to accomplish any given mission, avoid collisions, or even to maintain stability.

The use of aided navigation techniques, such as simultaneous localization and mapping (SLAM) algorithms, aims at solving this problem without using these, possibly compromised, sensing devices. SLAM is one of the most important parts of any autonomous system, and has been the subject of an extensive body of work by the research community. Aside from a myriad of variations on filtering techniques, applications, and sensors, there are three fundamental types of SLAM problems: range-and-bearing SLAM (RB-SLAM), range-only SLAM (RO-SLAM), and bearing-only SLAM (BO-SLAM). Landmark measurements for the latter sub-problems have a lower dimension than the considered mapping space, since a single noise-free observation provides only a line or surface as an estimate for the relative position of the landmark. Conversely, the more usual SLAM problem is sometimes referred to as range-and-bearing SLAM to underline the case where all the relative coordinates of measured landmarks are readily available. The interested reader can see [7] and [8] for a thorough survey on the algorithms proposed in the first decades of SLAM research, [9] for a specialized review of visual SLAM, [10] for a more up to date review focused in the recent theoretical achievements and [11] for an overarching survey on the history and remaining present and future challenges of SLAM, e.g., robustness and scalability, within others.

Being an intrinsically nonlinear problem, SLAM formulations tried to tackle two difficult theoretical challenges that are connected: algorithm inconsistency (see [12] and [13]) and convergence/stability (see [14], [15], [16] and [17]). Even though more involved strategies have since appeared, the robocentric framework proposed in [18] aimed at tackling the incon-

*Corresponding author

Email addresses: plourenco@isr.tecnico.ulisboa.pt (Pedro Lourenço), bj.guerreiro@fct.unl.pt (Bruno J. Guerreiro), pbatista@isr.tecnico.ulisboa.pt (Pedro Batista), paulo.j.oliveira@tecnico.ulisboa.pt (Paulo Oliveira), csilvestre@umac.mo (Carlos Silvestre)

sistency of EKF-SLAM while keeping the convergence problem open. It has also been used for computational gains, as in [19]. A natural evolution was to fully embrace the robocentric/sensor-based concept and free the filtering process of the pose of the vehicle. This was the approach considered in previous works by the authors, where globally asymptotically stable (GAS) sensor-based SLAM filters for 2-D and 3-D RB-SLAM were proposed, see [20] and [21] respectively. This trend was then applied successfully to RO-SLAM [22] and BO-SLAM [23] filters with exponential convergence (GES). These filters represent all landmark positions in the vehicle/sensor coordinate frame, making the positioning of the vehicle in the map trivial by construction (i.e. at the origin and aligned with the sensor frame) and consequently, not included in the state vector. The framework of the sensor-based SLAM filter is then completely independent of the inertial frame, as every input and state are expressed in the body-fixed frame. Nevertheless, most SLAM algorithms perform the mapping and localization in an inertial reference frame, as many applications require the inertial (or Earth-fixed) map and the trajectory of the vehicle. Examples of such applications are the fusion of maps resulting from different vehicles running separate sensor-based SLAM filters, or the same vehicle in different complementary scans of the environment, for which a consistent description of the maps and their uncertainty are of great relevance. For that reason, recovering the pose of the vehicle is important, and it is possible to do by matching the Earth-fixed map and the sensor-based map provided by the SLAM filter.

The authors have proposed a two-part strategy to tackle the problem of developing an online SLAM algorithm for unmanned aerial vehicles with global convergence properties. This strategy encompasses: (i) a sensor-based SLAM filter which estimates the landmark map and other vehicle related quantities expressed in the body-fixed frame; and (ii) an Earth-fixed Trajectory and Map (ETM) estimation algorithm resulting from an optimization problem with closed-form solution, which uses the sensor-based map estimate of the SLAM to provide a fully characterized uncertainty approximation for this highly nonlinear problem.

The approach presented in this paper generalizes the dual algorithm for the bidimensional RB-SLAM case proposed in [24], and the tridimensional RB-SLAM case first presented in [25]. In comparison with these preliminary conference works, which were examples of this idea applied to range-and-bearing sensor-based filters, this paper now presents: (i) the n-dimensional uncertainty characterization of the complete algorithm (which includes the computation of the new map as well as that of the pose); (ii) new simulations for the range-only formulation that explore the overall behaviour, as well as some of the decisions in the design of the algorithm; and (iii) results of new experiments for range-and-bearing SLAM in its bidimensional and tridimensional formulations, using a LiDAR and a Kinect camera respectively, as well as a larger experiment for bearing-only SLAM using a monocular camera and data from a widely available dataset, the *Rawseeds* dataset [26, 27]. In the proposed dual strategy, the pose of the vehicle can be estimated by matching the sensor-based and inertial maps, assuming that an initial Earth-fixed estimate is available. This counters the pure sensor-based framework, where the map obtained with the SLAM algorithm is expressed in the body-fixed frame. Therefore, in order to test the proposed algorithm, it is necessary to obtain sensor-based maps from other sources: the experiments presented herein complement previous purely sensor-based results in [20, 22, 21, 23].

The problem that underlies the algorithm proposed in this paper can be reduced to the computation of the transformation that maps two sets of points (Earth-fixed and sensor-based). This is usually called the Procrustes Problem [28]. Its generalization for rotation, translation and scaling has been subject of extensive research in areas such as computer vision applications, and can be traced back to [29] and [30]. The statistical characterization of this problem has also been the subject of study in works such as [28], [31], and [32]. However, some rather limiting options were taken, namely, the absence of weighting of the point sets, the use of small rotations, or the same covariance for all landmarks. This work proposes a methodology for obtaining the inertial map and the pose of the vehicle corresponding to a body-fixed map produced by a sensor-based SLAM filter, which builds on the n-dimensional formulation of the orthogonal Procrustes problem in [33] along with the thorough uncertainty characterization therein. It must be noted that [33] only deals with the equivalent of computing the pose in one particular instant, as opposed to an online computation of the pose along with the update of the inertial map. This is achieved resorting to perturbation theory, by considering arbitrary rotations and translations, individual weights, and individual covariance matrices for the landmarks of the inertial map. The results in this paper also hold for n-dimensions, which means that they can be applied directly to existing 2-D and 3-D algorithms, as shown by a variety of simulated and experimental trials herein reported.

The remainder of the paper is organized as follows. Section 2 presents an overview of the overall algorithm of which this work is an integral part, including the description of the generic sensor-based SLAM filter that complements this work. In Section 3, the ETM algorithm is described in detail, including the problem formulation, the uncertainty characterization, and the underlying Procrustes problem. Simulation results using a sensor-based RO-SLAM filter are presented in Section 4, and experimental results of three different implementations of sensor-based filters (2-D RB-SLAM, 3-D RB-SLAM, BO-SLAM) using different exteroception sensors are detailed in Section 5. Finally, concluding remarks and some directions for the future are presented in Section 6.

Notation. The superscript ^{E} indicates a vector or matrix expressed in the Earth-fixed frame $\{E\}$. For the sake of clarity, when no superscript is present, the vector is expressed in the body-fixed frame $\{B\}$. $\mathbf{I}_n$ is the identity matrix of dimension $n \times n$, and $\mathbf{0}_{n\times m}$ is a n by m matrix filled with zeros. If m is omitted, the matrix is square. $\mathbf{S}(\mathbf{a})$ is a special skew-symmetric matrix, henceforth called the cross-product matrix, as $\mathbf{S}(\mathbf{a})\,\mathbf{b} = \mathbf{a} \times \mathbf{b}$ with $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$. The skew operation is invertible, with $\mathbf{S}^{-1}(\mathbf{S}(\mathbf{a})) = \mathbf{a}$. The generalized anti-commutation matrix $\bar{\mathbf{S}}[\mathbf{a}] \in \mathbb{R}^{\frac{n(n-1)}{2}\times n}$ defined in [33, Definition 2] is also necessary. For $n = 3$ it reduces to $\mathbf{S}(\mathbf{a})$, and for $n = 2$ it is $\bar{\mathbf{S}}[\mathbf{a}] := \mathbf{a}^T\mathbf{S}(1)$. The matrix norm of a generic matrix $\mathbf{A}$ is defined as the Frobenius norm $\|\mathbf{A}\|^2 = \mathrm{tr}\left(\mathbf{A}\mathbf{A}^T\right)$, and the determinant is denoted as $\det(\mathbf{A})$ or $|\mathbf{A}|$. The expected value of any quantity is denoted by the symbol $\langle . \rangle$.

2. The complete methodology: an overview

The algorithm proposed in this paper was designed to complement a sensor-based SLAM filter, as part of a SLAM methodology that provides estimates in both the body-fixed frame and the local inertial/Earth-fixed frame. The final objective is to build a system that integrates a backend that performs the mapping in a relative frame, thus guaranteeing convergence and consistency, as well as a frontend that fixes that map to an Earth-fixed, or inertial, frame while providing the pose of the vehicle and the map expressed in the new coordinates. A diagram of the overall structure of the proposed SLAM methodology is presented in Fig. 1: (i) the exteroception sensor for each particular formulation provides data from which landmarks are detected and associated with the map. The particulars of the association process depend on the type of formulation and can be found in their respective papers [20, 21, 22, 23]; (ii) the measurement innovation vector is computed, and the observed/non-observed landmark sets are updated; (iii) the GES Kalman filter uses the proprioception data and the innovation to update the relative map and any vehicle variables (linear velocity, rate gyro bias, etc). Finally, (iv) this is then fed to the ETM algorithm proposed in this paper, which serves as the frontend and computes the pose and Earth-fixed map. A brief description of the sensor-based filter is provided in the next subsection, as an introduction to the ETM algorithm further detailed in Section 3. For an in-depth overview of the contents of this section, with expanded system design, observability results, and practical results, see [34].

2.1. Sensor-based SLAM

Building on the idea of robocentric filtering, [20] and [21] address the problem of designing a navigation system in a sensor-based framework for a vehicle capable of sensing the relative positions of landmarks in a previously unknown environment, in 2-D and 3-D respectively. This is done resorting to a purely sensor-based SLAM filter where no linearization or approximation is used whatsoever and pose representation in the state is suppressed, therefore avoiding its pitfalls. Due to the successful application of this strategy, further filters were designed in [22] and [23] for sensor suites measuring only ranges or bearings to landmarks. This subsection presents a brief overview of the filters proposed in those papers, which are shown to have globally asymptotically/exponentially stable error dynamics and serve as groundwork for the Earth-fixed Trajectory and Map estimation algorithm proposed in this work.

2.1.1. System design

Let $\{E\}$ denote the local inertial frame, hereafter referred to as the Earth-fixed frame, and $\{B\}$ denote the body-fixed frame, also known as the sensor-based frame, and $\mathbf{R}(t) \in SO(n)$, $n = 2, 3$, the rotation matrix that relates $\{B\}$ and $\{E\}$. This satisfies $\dot{\mathbf{R}}(t) = \mathbf{R}(t)\mathbf{S}(\omega(t))$, where $\omega(t) \in \mathbb{R}^{\frac{n(n-1)}{2}}$ is the angular velocity, expressed in body-fixed coordinates. Let also ${}^E\mathbf{p}(t) \in \mathbb{R}^n$ denote the position of the origin of $\{B\}$ described in $\{E\}$, and $\mathbf{v}(t) \in \mathbb{R}^n$ the velocity of the vehicle relative to $\{E\}$, expressed in $\{B\}$. The position and a map of the environment can be obtained using a SLAM algorithm, considering that the vehicle is equipped with: (i) a sensor that can measure information about the relative position, distance or direction, of landmarks, either naturally extracted from or artificially placed in the environment; as well as (ii) a triad of orthogonally mounted rate gyros; and, optionally, (iii) a sensor capable of measuring the linear velocity $\mathbf{v}(t)$.

In SLAM state-space formulation, the full state vector $\mathbf{x}(t) \in \mathbb{R}^{n_x}$, can be decomposed into vehicle specific variables dependent of the application, and landmark variables. These last contain the relative position and any further information regarding the corresponding landmark. Consider now that the N landmarks are divided in two different sets, depending on their visibility status: $\mathcal{M}_O := \{1, \ldots, N_O\}$ containing the N_O observed or visible landmarks and $\mathcal{M}_U := \{N_O + 1, \ldots, N\}$ containing the N_U unobserved, or non-visible, ones. Landmarks belonging to $\mathcal{M}_O$ will have some kind of system output associated, which leads to the definition of

$$\mathbf{y}_i(t) = \mathbf{f}(\mathbf{p}_i(t)), \quad i \in \mathcal{M}_O$$

where $\mathbf{y}_i(t)$ can be equal to $\mathbf{p}_i(t)$, $\|\mathbf{p}_i(t)\|$, or $\frac{\mathbf{p}_i(t)}{\|\mathbf{p}_i(t)\|}$, depending on whether the SLAM filter is based on range-and-bearing, range-only, or bearing-only measurements, respectively. Combining this information with the dynamics of the relative position of each landmark expressed in $\{B\}$, denoted as $\mathbf{p}_i(t)$, it is now possible to write the generic nonlinear system

$$\begin{cases} \dot{\mathbf{p}}_i(t) = -\mathbf{S}(\omega(t))\,\mathbf{p}_i(t) - \mathbf{v}(t) & i \in \mathcal{M} \\ \mathbf{y}_j(t) = \mathbf{f}(\mathbf{p}_j(t)) & j \in \mathcal{M}_O \end{cases}. \tag{1}$$

Note that this system implies that the landmarks $\mathbf{p}_i(t)$ are static in the Earth-fixed frame $\{E\}$, i.e.,

$${}^E\dot{\mathbf{p}}_i(t) = \mathbf{0}, \quad \forall i \in \mathcal{M}.$$

Depending on the chosen output equation, this system can be seen as linear time-varying (LTV), and, in the situations where it cannot, the authors in [22], [21], and [23] have transformed it with a suitable state augmentation and output transformation to yield an LTV-like system of the form

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t, \mathbf{y}(t), \mathbf{u}(t))\mathbf{x}(t) + \mathbf{B}(t, \mathbf{y}(t), \mathbf{u}(t))\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}(t, \mathbf{y}(t), \mathbf{u}(t))\mathbf{x}(t) \end{cases}. \tag{2}$$

where $\mathbf{x}(t) \in \mathbb{R}^{n_x}$ is the augmented state, $\mathbf{y}(t) \in \mathbb{R}^{n_y}$ is the (possibly) transformed output, $\mathbf{u}(t) \in \mathbb{R}^{n_u}$ is the system input, and $\mathbf{A}(t, \mathbf{y}(t), \mathbf{u}(t)) \in \mathbb{R}^{n_x \times n_x}$, $\mathbf{B}(t, \mathbf{y}(t), \mathbf{u}(t))\mathbb{R}^{n_x \times n_u}$, and $\mathbf{C}(t, \mathbf{y}(t), \mathbf{u}(t))\mathbb{R}^{n_y \times n_x}$ are the dynamics, input and output matrices that define the LTV system.

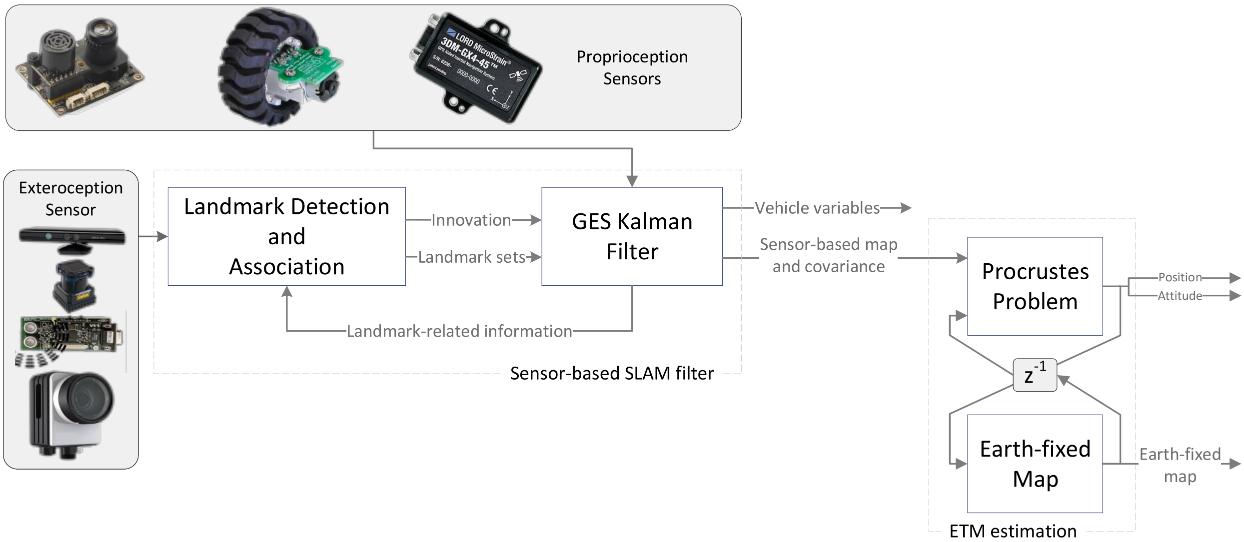


Figure 1: Diagram of the overall SLAM algorithm.

2.1.2. *Observer design*

In the referred papers, the observability analysis of each transformed system (2) is performed and conditions for the uniform complete observability are found. This analysis leads towards the design of a state observer, such as the Kalman filter for linear time-varying (LTV) systems, with globally exponentially stable error dynamics. It is also shown this observer converges exponentially fast to the true state of the nominal nonlinear system (1), thus effectively doubling as an observer for that system. The implementation of a SLAM filter for the resulting LTV system follows naturally with a linear discrete-time Kalman filter [35, 36], as all the sensors and the processing units are sample-based. Thus, the resulting discrete system is assumed to be perturbed with zero mean Gaussian noise, the measurements are assumed to be uncorrelated, and the estimates of the filter are characterized by their mean $\hat{\mathbf{x}}_k$ and covariance $\mathbf{\Sigma}_{\mathbf{x}_k}$, where $t_k = t_0 + k\,T_s$, T_s is the sampling time, $k \in \mathbb{N}_0$, and t_0 is the initial time.

2.2. *Problem Statement*

To complement the sensor-based filter, and due to the possible interest in obtaining Earth-fixed estimates for the vehicle pose and the environment map, there is a need for a strategy that takes the body-fixed map and the initial position and attitude as inputs and is able to compute for each time instant the current Earth-fixed map and the pose of the vehicle. For that purpose, the problem addressed in this paper is that of designing an algorithm to compute the Earth-fixed trajectory and map based on the information provided by a sensor-based SLAM filter as detailed in subsection 2.1. The idea is to write an optimization problem that can be related to the orthogonal Procrustes problem, and provide a characterization of the resulting uncertainty using perturbation theory.

3. Earth-fixed Trajectory and Mapping

This section formulates an optimization problem with a solution that corresponds to an estimate of the transformation between the body-fixed frame $\{B\}$ and the Earth-fixed reference frame $\{E\}$, yielding the algorithm here proposed. An error function is defined and then used to construct a cost function for the optimization problem. The algorithm builds on the derivation in [33] and the uncertainty characterization proposed therein, and aims at estimating, in real time, both the vehicle trajectory and the Earth-fixed map described in the same frame, with the respective uncertainty characterization.

3.1. *Formulation and Solution of the Problem*

The fundamental idea is to use the known sensor-based landmarks $\mathbf{p}_i$ to formulate an optimization problem with a closed-form solution to enable the real-time estimation of the transformation from $\{B\}$ to $\{E\}$ at each time instant k, which is defined by the position and orientation of the vehicle in $\{E\}$, respectively denoted as ${}^{E}\mathbf{p}_k \in \mathbb{R}^n$ and $\mathbf{R}_k \in \mathrm{SO}(n)$. This can be accomplished in several ways, however, the computation of the resulting transformation uncertainty depends highly on the choice of error that is minimized. Considering that the initial pose of the vehicle in $\{E\}$ is known *a priori* in any SLAM algorithm, as everything is then rooted back to that initial position and orientation. By design, the sensor-based filter provides a landmark map in the body-fixed frame, which means that the Earth-fixed map is readily available in the first instant. Assuming that the Earth-fixed landmarks considered for the ETM algorithm are static, as the vehicle navigates through the environment maintaining visibility with some of the previously visible sensor-based landmarks, it is always possible to compare the sensor-based map with that initial Earth-fixed map, which can be updated with new landmarks as they appear. This comparison fits precisely in the definition of the orthogonal Procrustes problem.

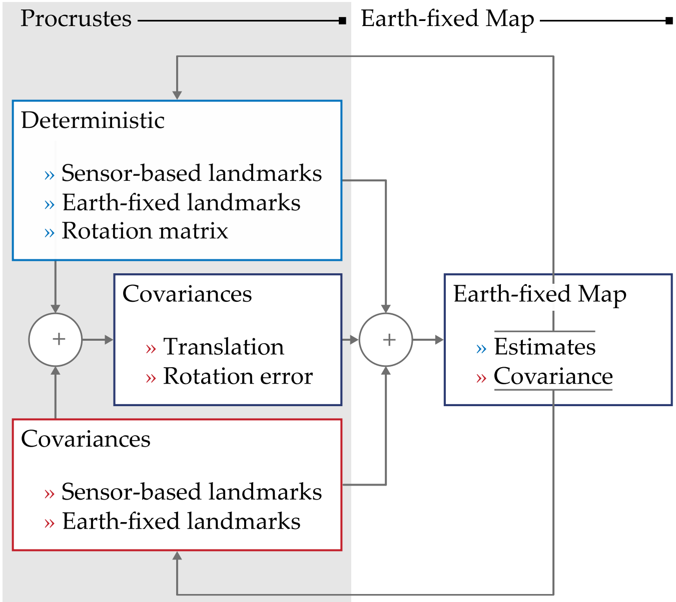


Figure 2: Summary of the work performed with the objective of designing the Earth-fixed Trajectory and Map estimation algorithm, showing the focus on the uncertainty of the Procrustes problem in [33] and the complete algorithm described in this paper.

Definition 1 (The orthogonal Procrustes Problem). *The problem of finding the translation and rotation that best describe the transformation between two related sets of points is called the Procrustes problem, named after a character in Greek mythology who made his victims fit his bed by either stretching their limbs or cutting them off [37].*

A closed-form solution to a particular formulation of this problem was first found in [38]. However, the uncertainty of that solution when the two sets are perturbed by noise is still a subject of research. In order to allow the use of this existing possibility in the problem at hand, it is necessary to have a measure of the resulting uncertainty. That is the focus of [33], where perturbation models consisting of deterministic and perturbed parcels are used to allow the computation of the covariances for the rotation and translation estimation errors as shown in Figure 2. Thus, [33] lays the foundation for the design of the Earth-fixed algorithm presented here, not only because it provides a method to compute the pose of the vehicle with its uncertainty characterization, but also because the methodology used therein can then be also applied to the step of updating the Earth-fixed map, allowing its uncertainty characterization as well. In summary, the proposed algorithm resorts to the solution of the orthogonal Procrustes problem presented in [33] to obtain the transformation between the sensor and Earth-fixed frames at each time instant, thus yielding the vehicle trajectory in the Earth-fixed frame. However, this implies that there is already a known Earth-fixed map of landmarks ${}^E\mathbf{p}_i$, $i \in \mathcal{M}$. With the static assumption, and having the sensor-based map, it is always possible to estimate the pose of the vehicle, using the Procrustes solution, and to initialize the Earth-fixed map by using the current vehicle pose estimate. Furthermore, it is also possible to update the current estimate of each landmark of the Earth-fixed map.

In general, at a given time instant k, if ${}^E\hat{\mathbf{p}}_k$ and $\hat{\mathbf{R}}_k$ are known, the update for each Earth-fixed landmark estimate can be computed using

$${}^E\hat{\mathbf{p}}_{i_k} = \hat{\mathbf{R}}_k \hat{\mathbf{p}}_{i_k} + {}^E\hat{\mathbf{p}}_k, \tag{3}$$

for all $i \in \{1, \ldots, N_M\}$. This algebraic loop may be averted if it is noticed that, as the landmarks in $\{E\}$ are static, there is also an equality between ${}^E\mathbf{p}_{i_{k-1}}$ and ${}^E\mathbf{p}_{i_k}$. This step is of the utmost importance in the design of the algorithm, and yields the following error function

$${}^E\mathbf{e}_{i_k} = {}^E\hat{\mathbf{p}}_{i_{k-1}} - \hat{\mathbf{R}}_k \hat{\mathbf{p}}_{i_k} - {}^E\hat{\mathbf{p}}_k, \tag{4}$$

for $k > 0$, that represents the error between the previous estimate of the Earth-fixed landmark and its sensor-based homologous at time k, rotated and translated with the estimated transformation. Naturally, dynamic landmarks cannot be dealt with by this strategy, but that is outside the scope of the algorithm. In fact, if dynamics landmarks are mapped, their dynamic behaviour will prevent future data associations and lead to their being eventually discarded. The pair $\left(\mathbf{R}_k, {}^E\mathbf{p}_k\right)$ can be obtained using the optimization problem

$$\left(\mathbf{R}_k^*, {}^E\mathbf{p}_k^*\right) = \arg \min_{\substack{\mathbf{R}_k \in SO(n) \\ {}^E\mathbf{p}_k \in \mathbb{R}^n}} G\left(\mathbf{R}_k, {}^E\mathbf{p}_k\right), \tag{5}$$

considering the cost function

$$G\left(\mathbf{R}_k, {}^E\mathbf{p}_k\right) = \frac{1}{N_T} \sum_{i=1}^{N_T} \sigma_{i_k}^{-2} \|{}^E\mathbf{e}_{i_k}\|^2.$$

The complete uncertainty present in the error function (4) is dependent on the uncertainty of the landmark estimate in each frame. However, in order to use the full uncertainty of the error function, the rotation from $\{B\}$ to $\{E\}$ would have to be known, which is not possible before solving the problem. For that reason, the uncertainty is approximated by the largest eigenvalue of each covariance summed, thus providing a conservative estimate of the uncertainty present in each pair $\left({}^E\mathbf{p}_{i_{k-1}}, \mathbf{p}_{i_k}\right)$. This is represented by $\sigma_{i_k}^2$, which is conservatively defined as

$$\begin{aligned}\sigma_{i_k}^2 &= \lambda_{max}(\mathbf{\Sigma}_{{}^E p_{i_k}}) + \lambda_{max}(\mathbf{\Sigma}_{p_{i_k}}) \\ &\geq \lambda_{max}(\mathbf{\Sigma}_{{}^E p_{i_k}} + \mathbf{R}_k \mathbf{\Sigma}_{p_{i_k}} \mathbf{R}_k^T).\end{aligned}$$

Note that the number of landmark pairs used in the ETM algorithm, N_T, may be different from the available number of landmarks, N, depending on the set of pairs of landmarks used in the optimization problem. It may be beneficial to use only a subset containing, for instance, the most recently visible (or least uncertain) landmarks in the sensor-based frame and their Earth-fixed analogous, provided that the dimension of the resulting sets is greater, if possible, than a predefined threshold. Other possibility is to choose the subset of sensor-based landmarks whose filtered estimates have already converged. Reducing the number of landmarks makes the algorithm more computationally efficient, and the threshold may be imposed to guarantee the numerical robustness and statistical consistency of the algorithm (see [33, Section 5] for more information on this). As mentioned before, the optimization problem (5) is in fact the orthogonal Procrustes problem as solved in [29] and [28], whose uncertainty characterization is proposed in [33]. From the original optimization problem it is known that the optimal translation is given by

$$\begin{aligned} {}^E\mathbf{p}_k^* &= \frac{1}{\sum_{j=1}^{N_T}\sigma_{j_k}^{-2}}\sum_{i=1}^{N_T}\sigma_{j_k}^{-2}\left({}^E\hat{\mathbf{p}}_{i_{k-1}} - \mathbf{R}_k^*\hat{\mathbf{p}}_{i_k}\right) \\ &= \boldsymbol{\mu}_{E_k} - \mathbf{R}_k^*\boldsymbol{\mu}_{B_k} \end{aligned} \tag{6}$$

which is the vector that translates the weighted centroid of the sensor-based landmarks $\boldsymbol{\mu}_{B_k}$ rotated to $\{E\}$ to the weighted centroid of the Earth-fixed landmarks $\boldsymbol{\mu}_{E_k}$. The optimal rotation is given by

$$\mathbf{R}_k^* = \mathbf{U}_k \operatorname{diag}\left(1, \cdots, 1, |\mathbf{U}_k||\mathbf{V}_k|\right)\mathbf{V}_k^T, \tag{7}$$

where $\mathbf{U}_k$ and $\mathbf{V}_k$ come from the singular value decomposition

$$\mathbf{U}_k\mathbf{D}_k\mathbf{V}_k^T = \operatorname{svd}\sum_{i=1}^{N_T}\sigma_{j_k}^{-2}\left({}^E\hat{\mathbf{p}}_{i_{k-1}} - \boldsymbol{\mu}_{E_k}\right)\left(\hat{\mathbf{p}}_{i_{k-1}} - \boldsymbol{\mu}_{B_k}\right)^T.$$

The Earth-fixed map estimate at instant k is computed using the update equation (3), following the computation of the optimal translation and rotation using the sensor-based map estimate of instant k and the Earth-fixed estimate of the previous iteration.

The work presented in this section, including the estimates for the vehicle pose, given by (6) and (7), and the update equation (3), allows the real-time computation of the vehicle trajectory and the Earth-fixed map. However, the ETM algorithm here described assumes the knowledge of the uncertainty of both the Earth-fixed and sensor-based landmark estimates. The latter is directly provided by the SLAM filter, but the former is yet to be described. The scope of this section is also to provide uncertainty descriptions of the estimates yielded by this algorithm, using perturbation theory and building on previous work proposed in [33].

3.2. Earth-fixed pose uncertainty characterization

The solution of the optimization problem (5) is based on the weighted orthogonal Procrustes problem. When dealing with sets of points stochastically perturbed as inputs, the result is itself perturbed and therefore it is important to look at the pair rotation-translation that is the solution along with a suitable uncertainty description. In [33] the authors studied thoroughly the underlying uncertainty and derived analytical expressions for the first and second moments of the stochastic outputs, the translation and rotation, as well as cross terms that characterize the anisotropic uncertainty of the problem, not imposing assumptions on the actual rotation and translation. For the problem at hands, this is done assuming the error models

$$\mathbf{p}_{i_k} = \mathbf{p}_{i_k}^{(0)} + \epsilon\,\mathbf{p}_{i_k}^{(1)} + O(\epsilon^2) \tag{8}$$

and

$${}^E\mathbf{p}_{i_k} = {}^E\mathbf{p}_{i_k}^{(0)} + \epsilon\,{}^E\mathbf{p}_{i_k}^{(1)} + O(\epsilon^2), \tag{9}$$

for the input landmark sets, while the translation and rotation follow the models

$${}^E\mathbf{p}_k^* = {}^E\mathbf{p}_k^{(0)} + \epsilon\,{}^E\mathbf{p}_k^{(1)} + O(\epsilon^2) \tag{10}$$

and

$$\mathbf{R}_k^* = \left(\mathbf{I}_n + \epsilon\,\mathbf{S}(\boldsymbol{\varpi}) + O(\epsilon^2)\right)\mathbf{R}_k^{(0)}, \tag{11}$$

where ϵ is the smallness parameter, the notation $O(\epsilon^m)$ stands for the remaining terms of order m or higher, $(.)^{(0)}$ are the zero order terms or true values, and the first order terms, $(.)^{(1)}$, are assumed to follow a known distribution with zero mean and covariance matrices defined by $\boldsymbol{\Sigma}_{(.)} = \langle(.)^{(1)}(.)^{(1)^T}\rangle$. The resulting covariances for the rotation error $\boldsymbol{\varpi} \in \mathbb{R}^{\frac{n(n-1)}{2}}$ and for ${}^E\mathbf{p}_k^*$ are thoroughly defined in [33, page 214], where they are denoted by $\boldsymbol{\Sigma}_\omega$ and $\boldsymbol{\Sigma}_t$. In this paper, however, they are henceforth presented as $\boldsymbol{\Sigma}_{\varpi_k}$ and $\boldsymbol{\Sigma}_{{}^Ep_k}$.

3.3. Earth-fixed map uncertainty characterization

The Earth-fixed map estimation depends on the translation and rotation estimates, as well as on a sensor-based map. Then, it is expected that the associated uncertainty will include terms related to the uncertainty of each one of these components, as well as cross covariance terms. Recall that the Earth-fixed map estimate is calculated with the update equation (3). Using this expression and the error models (8), (9), (10), and (11) it is possible to write

$${}^E\mathbf{p}_{i_k}^{(0)} = {}^E\mathbf{p}_k^{(0)} + \mathbf{R}_k^{(0)}\mathbf{p}_{i_k}^{(0)} \tag{12}$$

and

$${}^E\mathbf{p}_{i_k}^{(1)} = {}^E\mathbf{p}_k^{(1)} + \mathbf{S}(\boldsymbol{\varpi}_k)\,\mathbf{R}_k^{(0)}\mathbf{p}_{i_k}^{(0)} + \mathbf{R}_k^{(0)}\mathbf{p}_{i_k}^{(1)}. \tag{13}$$

Looking at (12), it is confirmed that ${}^E\mathbf{p}_{i_k}^{(0)}$ is the true quantity. Furthermore, from (13), it can be seen that ${}^E\mathbf{p}_{i_k}^{(1)}$ has zero mean, since all the quantities that compose it are themselves true quantities or have zero mean. The covariance matrix of the landmark position estimate is given by

$$\begin{aligned} \boldsymbol{\Sigma}_{{}^Ep_{ij_k}} &= \langle {}^E\mathbf{p}_{i_k}^{(1)}\,{}^E\mathbf{p}_{k_j}^{(1)^T}\rangle \\ &= \boldsymbol{\Sigma}_{{}^Ep_k} + \mathbf{R}_k^{(0)}\boldsymbol{\Sigma}_{p_{ij_k}}\mathbf{R}_k^{(0)^T} \\ &+ \langle {}^E\mathbf{p}_k^{(1)}\mathbf{p}_{j_k}^{(1)^T}\rangle\mathbf{R}_k^{(0)^T} + \mathbf{R}_k^{(0)}\langle\mathbf{p}_{i_k}^{(1)}\,{}^E\mathbf{p}_k^{(1)^T}\rangle \\ &+ \bar{\mathbf{S}}^T\left[\mathbf{R}_k^{(0)}\mathbf{p}_{i_k}^{(0)}\right]\boldsymbol{\Sigma}_{\varpi_k}\bar{\mathbf{S}}\left[\mathbf{R}_k^{(0)}\mathbf{p}_{j_k}^{(0)}\right] \\ &+ \bar{\mathbf{S}}^T\left[\mathbf{R}_k^{(0)}\mathbf{p}_{i_k}^{(0)}\right]\boldsymbol{\Sigma}_{\varpi_k{}^Ep_k} + \boldsymbol{\Sigma}_{\varpi_k{}^Ep_k}^T\bar{\mathbf{S}}\left[\mathbf{R}_k^{(0)}\mathbf{p}_{j_k}^{(0)}\right] \\ &+ \bar{\mathbf{S}}^T\left[\mathbf{R}_k^{(0)}\mathbf{p}_{i_k}^{(0)}\right]\langle\boldsymbol{\varpi}_k\mathbf{p}_{j_k}^{(1)^T}\rangle\mathbf{R}_k^{(0)^T} \\ &+ \mathbf{R}_k^{(0)}\langle\mathbf{p}_{j_k}^{(1)}\boldsymbol{\varpi}_k^T\rangle\bar{\mathbf{S}}\left[\mathbf{R}_k^{(0)}\mathbf{p}_{j_k}^{(0)}\right], \end{aligned}$$

where the covariances $\boldsymbol{\Sigma}_{^E p_k}$ and $\boldsymbol{\Sigma}_{\varpi_k}$, and the cross-covariance $\boldsymbol{\Sigma}_{\varpi_k {}^E p_k}$ are denoted as $\boldsymbol{\Sigma}_t$, $\boldsymbol{\Sigma}_\varpi$, and $\boldsymbol{\Sigma}_{\varpi t}$ in [33, pages 214-215]. The cross-covariances between the translation and the body-fixed map and between the rotation and the body-fixed map can be computed using $\langle \mathbf{t}^{(1)} {\mathbf{r}_j^{(1)}}^T \rangle \mathbf{R}_k^{(0)}$ and $\langle \boldsymbol{\omega}^{(1)} {\mathbf{r}_j^{(1)}}^T \rangle \mathbf{R}_k^{(0)}$, detailed in [33, Section D, supplementary material]. Finally, the cross-covariances between sensor-based landmarks $\mathbf{p}_{i_k}^{(1)}$ at time k and older Earth-fixed landmarks ${}^E\mathbf{p}_{i_{k-1}}^{(1)}$ for all i and j are assumed to be zero.

3.4. Implementation details

For the initialization of the Earth-fixed Trajectory and Mapping (ETM) algorithm, the traditional SLAM approach can be followed, by assuming that the transformation between Earth-fixed and sensor frames is known at time k_0, thus yielding naturally the landmark map in $\{E\}$ from the sensor-based map in $\{B\}$,

$$^E\hat{\mathbf{p}}_{i_0} = \mathbf{R}_0 \hat{\mathbf{p}}_{i_0} + {}^E\mathbf{p}_0,$$

with

$$\boldsymbol{\Sigma}_{^E p_{ij_0}} = \mathbf{R}_0 \boldsymbol{\Sigma}_{p_{ij_0}} \mathbf{R}_0^T.$$

Therefore, all the estimated Earth-fixed quantities are computed with respect to the initial pose. In a range-and-bearing formulation, where relative positions of landmarks are readily available on first observation, the use of this strategy is straightforward. However, for its use in conjunction with range-only or bearing-only filters, special care must be taken. Due to the fact that the algorithm uses the first observation as an Earth-fixed estimate, the algorithm can only be used when the uncertainty is low enough. Therefore, the first Earth-fixed pose estimate may not correspond to the beginning of the run. However, it is possible to use a GES smoothing filter to improve the initial sensor-based map estimates and then compute the missing trajectory.

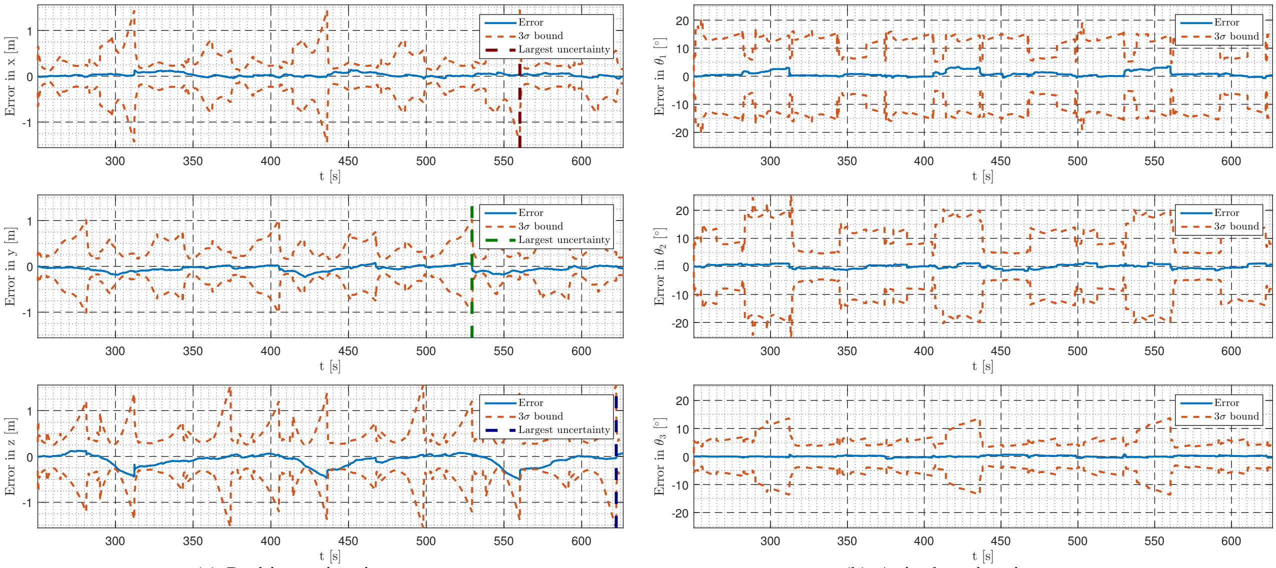
It is important to notice that in this procedure, an inertial landmark is only updated if the associated uncertainty decreases in that iteration. Thus, in each iteration, the candidate inertial landmarks covariance matrix is computed, and a measure of the uncertainty in each $\boldsymbol{\Sigma}_{^E p_{i_{k+1}}}$ is compared to its previous value. If the uncertainty is raised, then the old covariance is kept and the cross-covariances between an updated landmark and a non-updated one is set to zero, i.e., $\boldsymbol{\Sigma}_{^E p_{ij_{k+1}}} = \mathbf{0}$ for all $j \neq i$. There are many possible measures of the covariance, from the volume of the covariance ellipsoid (proportional to the determinant of $\boldsymbol{\Sigma}_{^E p_{i_k}}$) to the size of its largest axis (the maximum eigenvalue). In this work, the trace of $\boldsymbol{\Sigma}_{^E p_{i_k}}$ is used.

In the event that the attitude of the vehicle is provided by an external source, such as an AHRS [39], then only the Earth-fixed position and map need to be estimated using (6) and (3). In that case all the cross-terms involving $\boldsymbol{\varpi}_k$ are zero and $\boldsymbol{\Sigma}_{\varpi_k}$ must be either estimated or provided by the AHRS. Furthermore, if a relative map is obtained in a world-centric framework, as in [40], the position and Earth-fixed map can still be found as before, and then all the occurrences of $\mathbf{R}_k^{(0)} \mathbf{p}_{i_k}^{(\cdot)}$ are substituted by the correspondent quantity in the world-centric relative map and all quantities involving $\boldsymbol{\varpi}_k$ disappear.

4. Simulation results: Range-only SLAM

The ETM strategy proposed in this paper is completely independent of the source of the sensor-based map that serves as input. This means that, as long as it is possible to extract a coherent map using range and/or bearing measurements, the ETM algorithm will be able to match that evolving map to a frame fixed to the initial pose. However, depending on the source there are some particularities that influence the result. For example, in range-and-bearing or bearing-only procedures where landmark association is not perfect, the Earth-fixed pose and map obtained may be impaired by erroneous associations.

To show a situation where this does not occur, consider the sensor-based range-only SLAM filter proposed by the authors in [22]. This filter uses distances to landmarks and measurements of linear and angular velocities, yielding as output the body-fixed map (along with other auxiliary quantities). It should be noted that the association of landmarks is known for range-only frameworks, since each beacon that communicates with the on-board transceiver sends an identified acoustic/electromagnetic signal. Therefore, it is possible to tag each of the landmarks in the map. A simulated environment was devised in Matlab to better explore several aspects of the algorithm and showcase its performance (a complete description of this environment can be found in [41, Appendix C]). This tries to emulate the fifth floor of the North Tower at IST. It consists of a 16 by 16 by 3 m corridor, where 36 landmarks were put in notable places such as corners and doors, with random heights. Since landmarks in range-only SLAM are not observable for all trajectories, the simulated trajectory is designed to excite the system in all directions to guarantee observability of the map (see [22] for further details). The aerial vehicle is initialized on the ground and, after take-off, makes several laps around the corridor, as shown in Figure 6. The total distance travelled is 294 meters in 627 seconds, at an average speed of 0.469 m/s, while one single loop of 58 m along the corridor takes 124 s. All the measurements are assumed to be perturbed by zero-mean Gaussian white noise, with standard deviations of $\sigma_\omega = 0.05$ °/s for the angular rates, $\sigma_v = 0.03$ m/s for the linear velocity, and $\sigma_r = 0.03$ m for the ranges. Taking advantage of the simulated environment, several runs are performed with different decision factors for the update step of the algorithm: the trace of the covariance of the individual landmark, the determinant, the largest eigenvalue (all with comparable accuracy in the resulting estimates). An additional run with updates every time-step is also provided. Statistics for the results of estimating position, attitude and maps for these four variants are presented in Tables 1, 2, and 3. There the mean, standard deviation, and root mean square of the absolute trajectory (A.2), absolute attitude (A.3), and relative pose errors (A.4) (see Appendix AppendixA for the definitions of each of these errors) are presented for the three products of the algorithm: position, attitude, and map. The statistics of a single landmark (see Figure 4) are also shown in the *"1 Landmark"* column of Table 3. To allow a better understanding of the performance of the ETM algorithm, the results for the sensor-based quantities fed to it are also presented: average landmark error norm and its



(a) Position estimation error. (b) Attitude estimation error.

Figure 3: RO-SLAM: Evolution of the vehicle pose estimation error with uncertainty bounds.

Table 1: Errors for the RO-SLAM simulation results (mean±standard deviation · RMSE), with the different ETM update criteria.

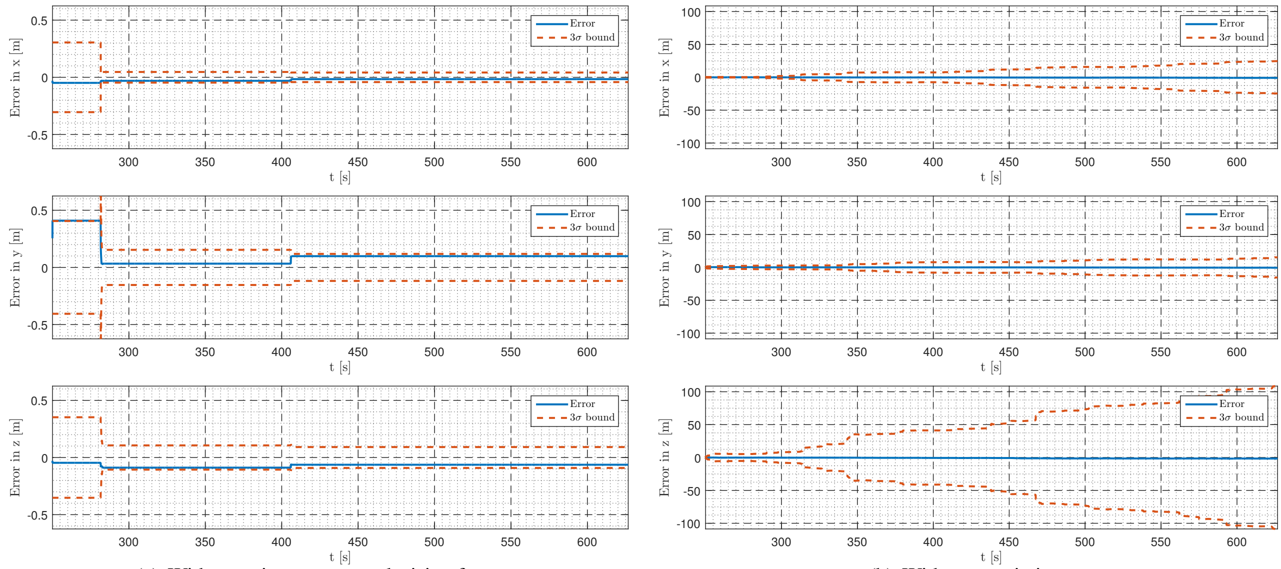
Alg.	Position (ATE) [m]	Attitude (AAE) [deg]
ETM, trace decrease	0.142±0.115 · 0.183	1.273±0.806 · 1.507
ETM, determinant decrease	0.133±0.091 · 0.161	1.183±0.619 · 1.335
ETM, $\lambda_{\max}$ decrease	0.151±0.115 · 0.189	1.211±0.844 · 1.476
ETM, unrestricted updates	0.511±0.334 · 0.611	4.679±2.549 · 5.328

Table 2: Pose errors (RPE) for the RO-SLAM simulation results (mean±standard deviation · RMSE), with the different ETM update criteria.

Alg.	Δ = 1 frame [m]	Δ = 1 sec [m]	Δ = 1 min [m]
Trace	0.002±0.008 · 0.008	0.016±0.024 · 0.029	0.258±0.134 · 0.291
Det.	0.002±0.007 · 0.007	0.015±0.020 · 0.025	0.242±0.122 · 0.271
$\lambda_{\max}$	0.002±0.008 · 0.008	0.016±0.024 · 0.029	0.255±0.135 · 0.288
Unrest.	0.002±0.001 · 0.002	0.012±0.008 · 0.014	0.225±0.111 · 0.251

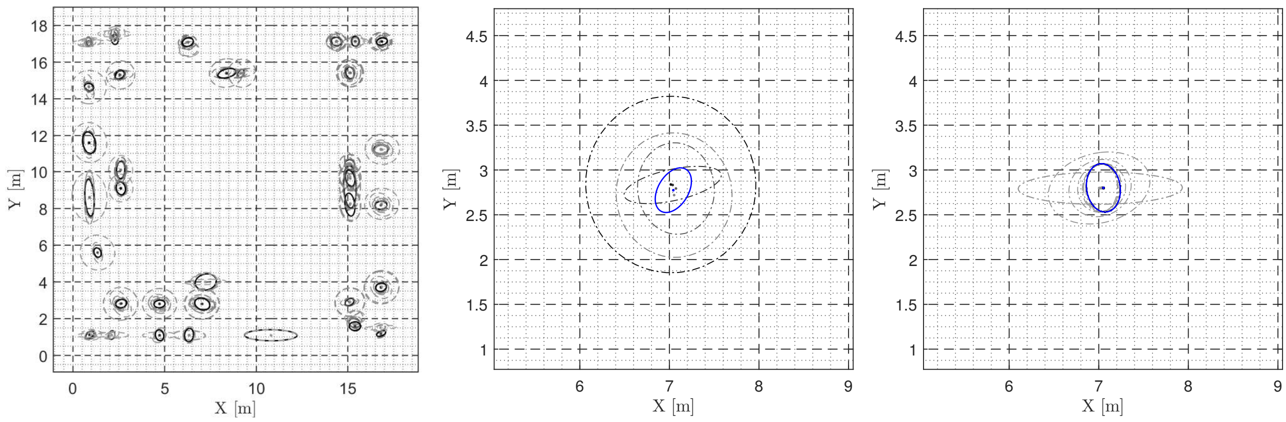
Table 3: Absolute map errors for the RO-SLAM simulation results (mean±standard deviation · RMSE). On top, the sensor-based RO-SLAM results, and below the ETM results with the different update criteria.

Algorithm	1 Landmark (ALE) [m]	Map (AME) [m]
Sensor-based RO-SLAM	0.296±0.457 · 0.544	1.145±6.172 · 6.277
ETM, trace decrease	0.029±0.026 · 0.039	0.171±0.137 · 0.219
ETM, determinant decrease	0.058±0.001 · 0.058	0.170±0.132 · 0.215
ETM, $\lambda_{\max}$ decrease	0.053±0.009 · 0.054	0.179±0.141 · 0.228
ETM, unrestricted updates	0.338±0.397 · 0.521	0.583±0.391 · 0.701



(a) With covariance trace as decision factor. (b) Without restrictions.

Figure 4: RO-SLAM: One particular landmark position error and uncertainty, with two different runs of the algorithm.



(a) Full map. (b) Detail of one sensor-based landmark. (c) Detail of one Earth-fixed landmark.

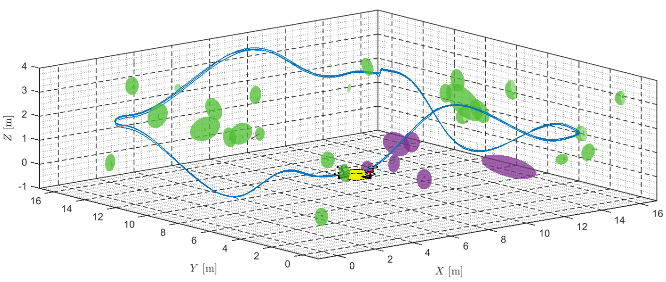
Figure 5: RO-SLAM: The spatial evolution of the map covariances. On the left, overlapped snapshots of the Earth-fixed map taken every second. On the center, details of the evolution of the input sensor-based covariance for one of the landmarks in the map, taken every 100 seconds and rotated to the Earth-fixed frame. On the right, the corresponding output Earth-fixed covariance every second, respectively. In this figure, the darker the colour of the ellipse the later in the run it appears.

standard deviation for the same single landmark and the whole map.

Figure 3(a) shows the evolution of the vehicle position estimation error for a typical simulation in solid blue and the accompanying 3σ uncertainty bound. It can be seen that the error is kept at a low level and that the algorithm manages to keep consistent estimates throughout the run. This assertion is confirmed by Figure 3(b), which shows the attitude estimation error in solid blue and the 3σ uncertainty bounds. Since the attitude estimates provided by ETM are in the form of a rotation matrix, the attitude estimation error is immediately accessible as explained in (A.1). This, as the position error, is always low and inside the confidence interval provided by the uncertainty characterization. Note that the uncertainty characterization depends solely on the sensor-based map and the vehicle attitude. Since the trajectory contains periodic repetitions of the pose of the vehicle, and consequent loop closings occur constantly throughout the run, the computed uncertainty shows also a periodic nature. However, there are clear peaks in the position uncertainty, identified for each direction with a vertical dashed line. The poses corresponding to these peaks are depicted in Figure 6(a) following the same colour code.

To understand the impact of restricting updates to uncertainty decreases, consider Figure 4. On the left side (Figure 4(a)), the evolution of the error and corresponding uncertainty when the decision to update is based on the decrease of the trace of the covariance, and on the right side (Figure 4(b)) the results for unrestricted updates. It can be seen that not requiring an uncertainty decrease renders the estimates unstable since the uncertainty has a rising trend with local decreases. This is confirmed in the average results shown in Table 1, where the errors are considerably larger and with higher variance than in the restricted versions. On the other hand, using the covariance trace as the factor provides a fast convergence in all directions, as do the other restricted versions. Furthermore, the overall map error throughout the run is decreased from the sensor-based estimates to the Earth-fixed ones, which is an interesting and expected effect. Indeed, the inertial model is exact, as the landmarks remain constant, which allows for its uncertainty to decrease over time as more measurements are obtained.

(a) Top-view of the map.



(b) 3-D map.

Figure 6: RO-SLAM: Earth-fixed map and trajectory with uncertainty bounds for both landmarks and equally time-spaced poses. Top-view of the map with the floor blueprint in the background.

To provide a general overview of the results of the ETM algorithm, Figure 6 shows the estimated trajectory and map at the end of the run. The small quadrotor indicates the pose at the current moment, the blue line is the estimated path, the red dashed line is the ground truth, the purple (green) ellipses/ellipsoids are the landmarks visible (occluded) at the current time. The ellipses along the path represent the 3σ bounds for the pose estimates at that point. The black circumferences and red circles are the real landmark and vehicle positions, respectively. These figures show that the map and the estimated trajectory are quite accurate and the uncertainty characterization is consistent with the real level of error displayed.

Finally, Figure 5 shows the spatial evolution of the uncertainty of the whole Earth-fixed map and a detail comparing the varying uncertainty of a sensor-based landmark, which rotates, increases and decreases throughout the run, and the corresponding uncertainty of the Earth-fixed estimate. The ellipses from covariances estimated by the ETM procedure (Figures 5(a) and 5(c)) are taken every second and help visualize the convergence of the uncertainty and the shapes it assumes throughout time. To better understand the behaviour of the procedure, the uncertainty ellipses for the same sensor-based landmark are also presented in Figure 5(b), where the various ellipses were obtained with intervals of 100 seconds and transformed to the Earth-fixed frame for a fair comparison. In all of these sub-figures of Figure 5, the darker the colour of the ellipse the later in the run it appears.

5. Experimental results

This section provides experimental results for the performance and consistency evaluation of the ETM algorithm coupled with the sensor-based SLAM filters in [20], [21] and [23]. Since the ETM algorithm works in cascade after the sensor-based filter, its performance is obviously dependent on the individual performance of the latter. The separation of the estimation process in these two stages aims at providing a less uncertain and consistent Earth-fixed trajectory and landmark map. This assertion is grounded on the notion of performing control, decision, and loop-closing procedures in the sensor-based frame, lessening possible effects of nonlinearities in the filter. However, the final Earth-fixed result is still affected by the nonlinearity inherent to the problem of arbitrarily transforming maps between coordinate frames, which is common in EKF-based SLAM algorithms. It should be noted, that, while the uncertainty characterization is approximate, it is usually possible to have always more than 10 landmark pairs in ETM iterations, hence guaranteeing that the uncertainty characterization is valid, as shown by the extensive tests in [33]. In summary, the results reported in this section serve as practical validation of the approach proposed in this paper as a whole and, while still an indication of the performance of its integral parts, these have to be viewed together. Those that include position and attitude ground truth are summarized in Tables 4, 5, and 6, and Figure 7 provides a graphic overview of each experimental setup. An in-depth report is provided in the remainder of the section.

5.1. Range-and-bearing SLAM

The first of the experiments in this paper combines a sensor-based SLAM filter for range-and-bearing measurements with the ETM algorithm. Three runs are detailed: (i) a 60 m long 2-D experiment using a LiDAR without ground truth [20] for which a video can be found in [42]; (ii) a 20 m long 3-D experiment using an RGB-D camera with ground truth; and (iii) an 80 m long 3-D experiment using an RGB-D camera with partial ground truth (both detailed in [21]) for which a video can

Table 4: The sensor suites for all the experiments documented in this section.

Quantities	Sensors
Landmark position	LiDAR (RB-2D) / RGB-D camera (RB-3D)
Landmark bearing	Monocular camera (BO)
Linear velocity	Odometry (BO) / Estimated by the SLAM filter (RB-2D,RB-3D)
Angular velocity	IMU (RB-2D,RB-3D,BO)
Ground truth	No GT (RB-2D) / VICON Motion capture (RB-3D) / Industrial cameras + laser scan matching (BO)

Table 5: Errors for the experimental results (mean±standard deviation · RMSE)

Algorithm	Position (ATE) [m]	Attitude (AAE) [deg]
RB-SLAM (Run 3-D, #1)	0.097±0.076 · 0.124	3.695±3.673 · 5.209
BO-SLAM	1.220±0.494 · 1.316	0.429±1.074 · 1.156

be found in [42]. The sensor suites of the RB-SLAM filters that feed the ETM algorithm in these experiments are composed of only two sensors each: a LiDAR or RGB-D camera to measure relative positions of landmarks and rate-gyros for the angular velocity. The filters output the map, the linear velocity and the rate-gyro biases, all expressed in body-fixed coordinates.

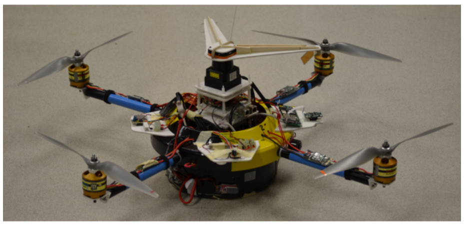
5.1.1. 2-D RB-SLAM results

An instrumented quadrotor was hand-driven along a path of about 60 meters in an indoor environment with a loop, as shown in Figure 8, at an average speed of 0.4 m/s. The trajectory described by the vehicle starts near the middle and moves counterclockwise until some of the first landmarks detected are once again visible, at the lower right corner. This custom quadrotor UAV, property of ISR, is equipped with a *MEMSENS nanoIMU*, a *Maxbotix XL* sonar for altitude measurements, and a *Hokuyo UTM-30LX* laser scanning device that provides horizontal profiles of the surroundings. These are fed to a landmark detection algorithm, where the data is processed into clusters and a robust line detection strategy is implemented, which builds on the basic split and merge algorithm, see [43], using a reduced space Hough transform parameter fitting for each line, in a similar way to the algorithm proposed in [44]. Afterwards, a corner identification procedure is used to obtain the desired point landmarks. Further details about the experimental setup regarding the sensor-based SLAM filter can be found in [20].

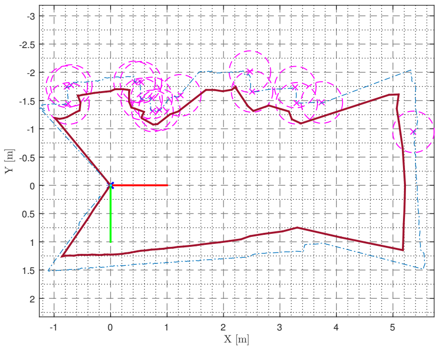
Besides the optimization-based Earth-fixed trajectory and map (ETM) estimation algorithm proposed in this paper, these experimental results also describe the traditional approach of augmenting the sensor-based SLAM filter with the pose of the Earth-fixed frame, hereafter denoted as the AugETM algorithm. This is done by including the initial position in the filter state as an artificial landmark and the orientation represented as the yaw angle, as described in more detail in [24]. With this simple implementation, it is possible to filter the Earth-fixed position as an (un-observable) landmark in the map. This functions as a kind of dead-reckoning, where the position and orientation are obtained by integration of the linear and angular velocities (and their associated uncertainties), with the difference that the former is not measured but estimated by the SLAM filter. Therefore, there is a significant difference between the two algorithms, as the uncertainty in the AugETM always grows, contrary to the ETM that reflects the combined uncertainty of the two maps.

Using either one of these methods, the trajectory of the vehicle and a consistent map in the Earth-fixed frame can be obtained. The Earth-fixed maps computed using both the ETM and AugETM algorithms are presented in Figure 10, featuring the final results of the experimental trial. The landmarks and their respective 95% confidence bounds are shown in three different styles (solid magenta, dashed yellow, and dash-dotted light blue), denoting respectively the three sets of landmarks used for the loop closing step (as described in [20]): i) recently visible landmarks, ii) old landmarks, and iii) intermediary landmarks. For the Earth-fixed maps, the trajectory of the vehicle and the respective 95% confidence bounds are shown (in dash-dotted green for ETM and in dashed dark blue for AugETM), whereas the laser sensor readings are shown in solid light gray. For completeness, the Earth-fixed coordinate frame, which was added to the filter as an unobservable landmark, is shown in the sensor-based map (axes shown in red and green, position and uncertainty in dark blue), along with the vehicle trajectories computed using both ETM and AugETM, respectively, in green and dark blue. The main aspects to retain from these intricate figures are the quality of the Earth-fixed map generated by the algorithm and the smaller landmark uncertainty of the Earth-fixed map generated using the ETM, when compared with the one generated using the AugETM approach.

In the sensor-based framework, as there is no vehicle localization uncertainty and the full map rotates and translates when the vehicle moves, the uncertainty of the non-visible landmarks increases with time, thus enabling a consistent sensor-based map estimation. Conversely, in the Earth-fixed map the landmarks are assumed static and, therefore, the landmark position uncertainty is always non-increasing, whereas the vehicle pose uncertainty may increase indefinitely unless a loop closure scenario occurs. The time evolution of the orientation and position uncertainties, using both ETM and AugETM approaches, are presented in Figure 9. It can be seen that the pose uncertainty computed with AugETM increases even when the vehi-

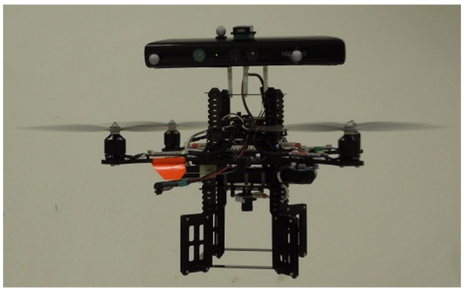


(a) RB-SLAM: Run 2-D – Instrumented quadrotor for data acquisition.

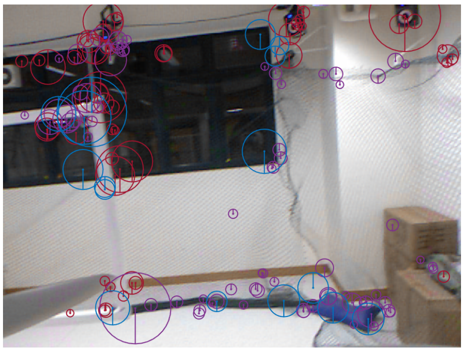


(b) RB-SLAM: Run 2-D – Laser profile with detected features.

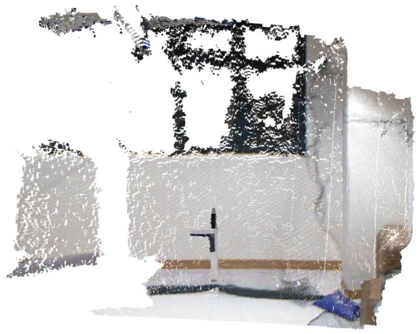
(c) RB-SLAM: Run 2-D – Picture of the region in Figure 7(b).



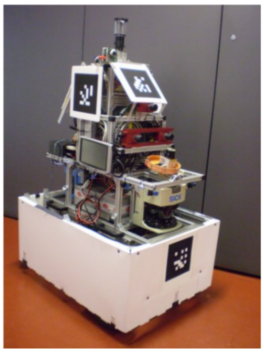
(d) RB-SLAM: Run 3-D – Instrumented *AscTec Pelican* quadrotor for data acquisition.



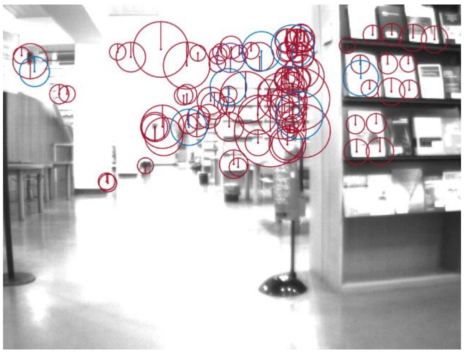
(e) RB-SLAM: Run 3-D – *Kinect* RGB picture with detected features.



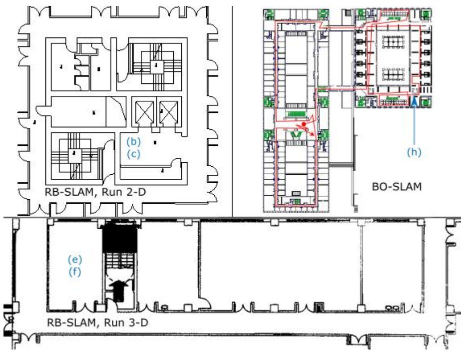
(f) RB-SLAM: Run 3-D – *Kinect* pointcloud corresponding to Figure 7(e).



(g) BO-SLAM: Instrumented ground robot used in the experimentis in the *Rawseeds* dataset ([26] and [27]).



(h) BO-SLAM: Picture of a typical space with detected features.



(i) The blueprints of the regions covered in all trials.

Figure 7: Information on the experimental setup of each trial described in this section. From top to bottom: RB-SLAM in 2-D, RB-SLAM in 3-D, BO-SLAM. On the left, the data acquisition platforms for each algorithm. On the bottom right, the locations of the acquisition of each picture/LiDAR scan.

Table 6: Pose errors (RPE) for the experimental results (mean±standard deviation · RMSE)

Algorithm	Δ = 1 frame [m]	Δ = 1 sec [m]	Δ = 1 min [m]
RB-SLAM (Run 3-D, #1)	0.004±0.007 · 0.008	0.070±0.062 · 0.093	0.220±0.132 · 0.256
BO-SLAM	0.029±0.028 · 0.041	0.044±0.032 · 0.055	0.344±0.138 · 0.371

cle is immobilized (first 50 seconds), whereas the pose uncertainty computed by the ETM algorithm remains constant. The standard state augmentation method is based on the open-loop integration of the state of the sensor-based filter, noting that the velocity disturbance noise, necessary for the consistency of the sensor-based filter, is also integrated to obtain the Earth-fixed frame position. Conversely, the ETM algorithm uses the best landmark position estimates in both $\{B\}$ and $\{E\}$ frames to com-

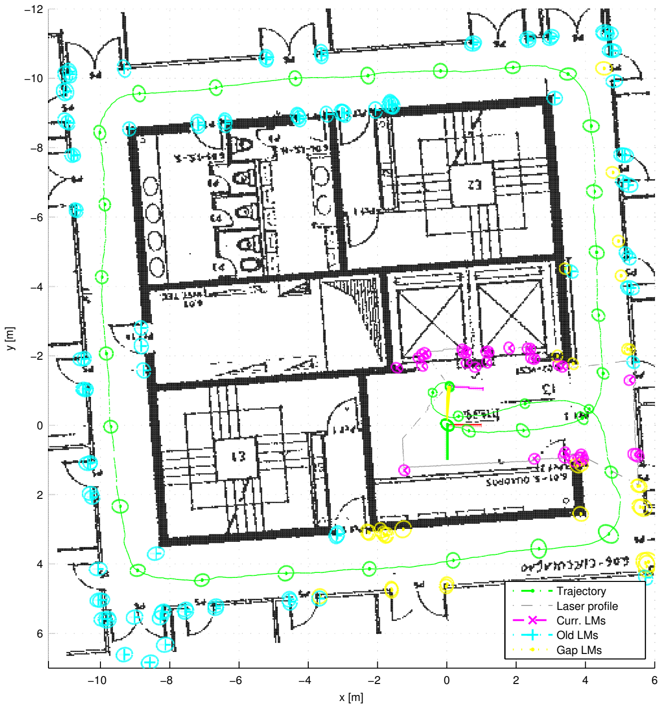
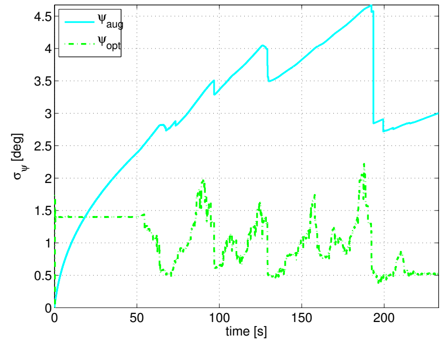
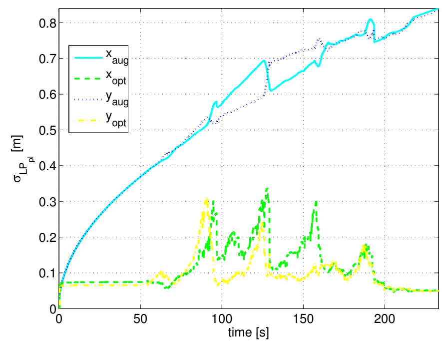


Figure 8: RB-SLAM: Run 2-D – Map and trajectory in the Earth-fixed frame, obtained using the optimization-based Earth-fixed trajectory and map algorithm, with the floor blueprint on the background.



(a) Heading Uncertainty



(b) Position Uncertainty

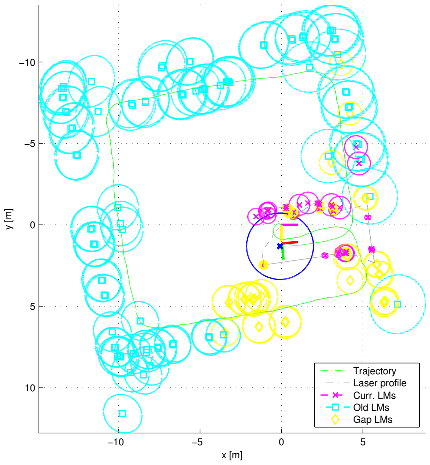
Figure 9: RB-SLAM: Run 2-D – Vehicle pose uncertainty in the Earth-fixed frame, using the augmented state approach AugETM (in solid light blue and dotted dark blue) and using the optimization-based approach ETM (in dashed green and dash-dotted yellow).

pute new estimates of the Earth-fixed map and vehicle pose. The result is notorious when comparing the Earth-fixed landmark uncertainties arising from each method in Figs. 10(b) and 10(c). The heading of the vehicle is also shown in Figure 11, where it can be seen that both transformations provide coherent results and, notably, that the effects of the magnetic distortions inside buildings can be devastating. Note that the heading solution provided by the external attitude filter is utterly wrong, which is a consequence of relying on the IMU magnetometer readings while indoors.

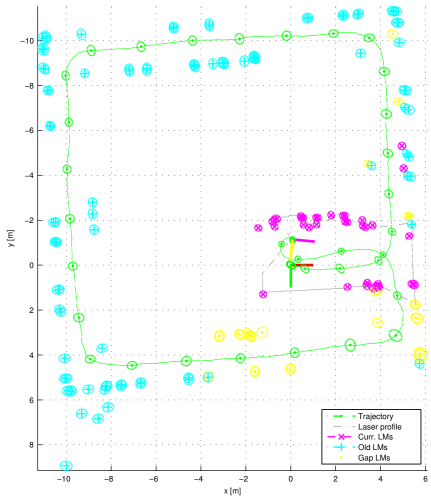
In the experimental results presented in [20], the moments just before and right after the loop closure are shown for the sensor-based map. Loop closings happen when associations between landmarks in the old and recent sets are found, after searching periodically using an implementation of the JCBB algorithm [45]. Even though the loop closing occurs in the sensor space, using only information from the sensor-based SLAM filter, this transition instant is also presented for the Earth-fixed map generated for the ETM algorithm, in Figure 12, where a detailed version of the map of Figure 10(b), for the first loop closing instant, is presented. The landmark associations between the current and old sets are shown in solid black and the fused landmarks positions and uncertainty bounds obtained after the loop closure are also depicted in solid black. These results indicate that the Earth-fixed map provided by the proposed ETM algorithm is also consistent.

5.1.2. 3-D RB-SLAM results

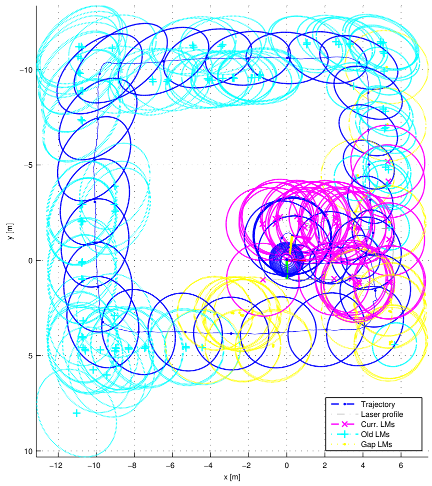
The above results were complemented for three dimensions by a series of experiments conducted in the Sensor-based Cooperative Robotics Research Laboratory - SCORE Lab of the Faculty of Science and Technology of the University of Macau. The setup consists of an *Asctec Pelican* quadrotor equipped with a *Microstrain 3DM-GX3-25* inertial measurement unit working at 50 Hz and a *Microsoft Kinect* camera, at 10 Hz (see [41, Appendix D] for further details on the experimental setup.).The experiments consisted in moving the quadrotor inside a 8m×6m room (usable area of 16 m^2) equipped with a *VICON* motion capture system [46, 47], a state-of-the-art optical system that records the movement of objects or people with millimetric resolution in 3-D, by means of infrared marker-tracking. This system provides accurate estimates of the position, attitude, linear and angular velocities of any vehicle placed inside the working area with the correct markers. Aside from the fact that a full tridimensional approach is used here, there are several aspects of the results herein described that differentiate them from the 2-D case, as the process by which landmarks are obtained and the existence of ground truth to validate the results. In summary, the sensor-based SLAM filter used for these experiments relies on RGB-D (colour and depth) images acquired by the *Kinect*. The RGB images are fed to a SURF implementation [48], which detects a 64-dimensional descriptor of features on the 2-D pictures of the environment. These



(a) Sensor-based frame, $\{B\}$.



(b) Earth-fixed frame, $\{E\}$, using ETM algorithm.



(c) Earth-fixed frame, $\{E\}$, using AugETM algorithm.

Figure 10: RB-SLAM: Run 2-D – Maps and trajectories in the sensor and Earth-fixed frames. Note that the y-axis scale must be inverted so that the frames used in the algorithm are correctly defined, since the vehicle uses the north-east-down convention.

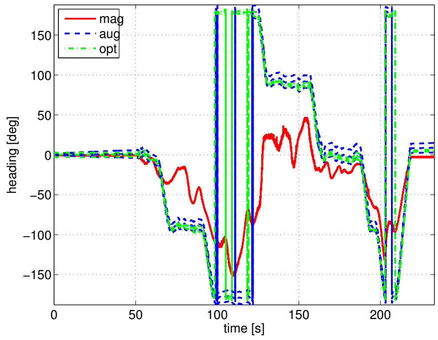
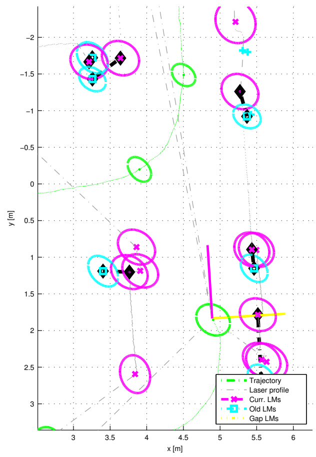
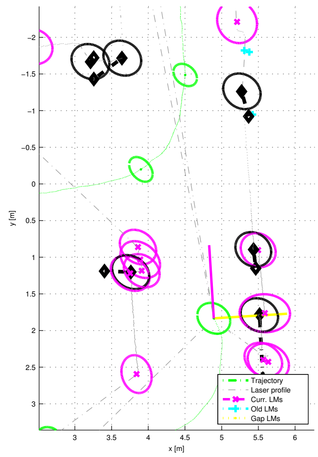


Figure 11: RB-SLAM: Run 2-D – Heading of the vehicle relative to the Earth-fixed frame, with 2σ confidence bounds, provided by: an external attitude filter based on magnetometer readings (solid red); the augmented state approach AugETM (dashed dark blue); and the optimization approach ETM (dash-dotted green).



(a) Before loop closure.



(b) After loop closure.

Figure 12: RB-SLAM: Run 2-D – Detailed view of the map and trajectory in $\{E\}$, using ETM algorithm, before and after the loop closing procedure.

features are then matched to a pointcloud built with the depth image. A greedy association process based on the Mahalanobis distance and the 64-D descriptor is used to associate new measurements with existing data. Technical details on the actual implementation can be found in [21]

The results of two different experiments are detailed here. In the first experiment, depicted on Figures 13-15, the vehicle does not leave the area covered by the *VICON* system and, as such, ground truth is always available. The second, longer, experiment consists of a small lap inside the lab followed by a larger exploration of the outside corridor as shown in Figure 16. In both runs, the vehicle is hand-driven at an average speed of around 0.4 m/s. Figures 13 and 14 depict the position and orientation estimates against the ground truth and 95% uncertainty bounds for the first run. It can be seen that the estimates are rather close to the ground truth, and, although there are some moments where the algorithm is somewhat optimistic, the overall performance shows consistency. However, the vertical performance is worse than the horizontal one, which is

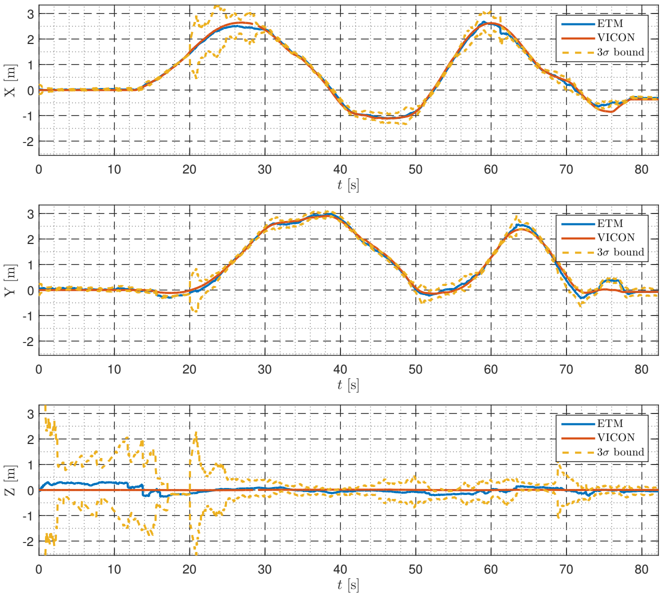


Figure 13: RB-SLAM: Run 3-D, #1 – Evolution of the position estimates with ground truth and uncertainty bounds. Horizontal trajectory (top two figures), Vertical trajectory (bottom).

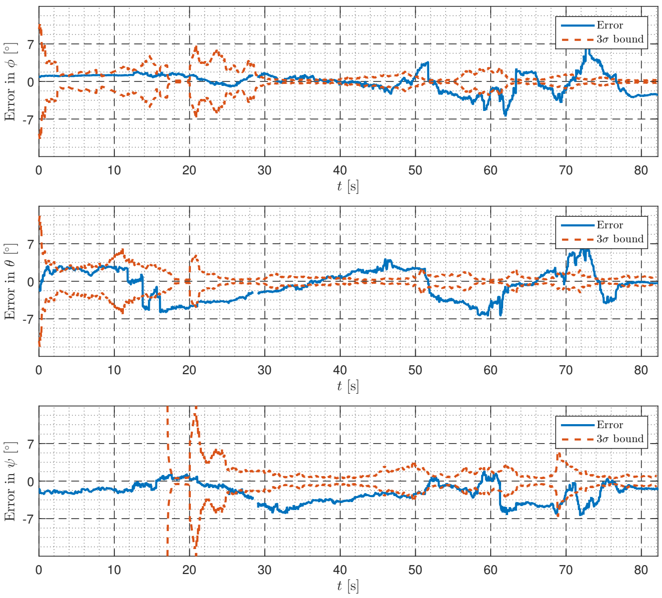


Figure 14: RB-SLAM: Run 3-D, #1 – Evolution of the estimation error of the Euler angles with uncertainty bounds. From top to bottom: roll, pitch, and yaw.

quite accurate. This is further apparent in Figure 14, where the pitch and roll are shown to be estimated with higher relative error. This is most likely due to the lack of vertical motion by the vehicle (the trajectory is mostly two-dimensional), and the reduced angle-of-view of the *Kinect* camera which limits the vertical separation of landmarks (smaller baseline). Note that for this filter there is no odometry, and all the linear information is being extracted by the association of landmarks detected in the *Kinect* pointclouds. Furthermore, this data is also correcting online the bias and any drift introduced by the rate gyros.

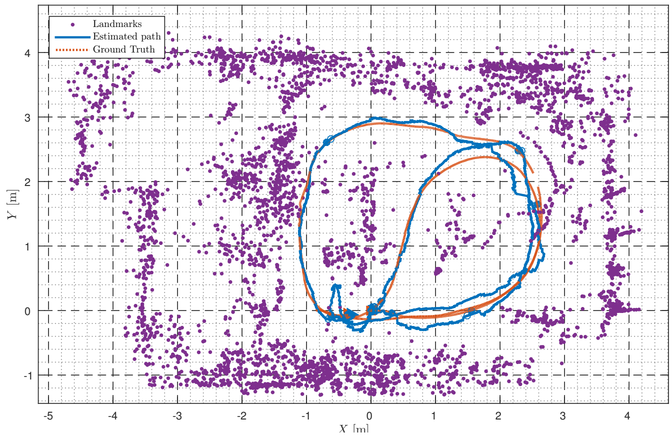


Figure 15: RB-SLAM: Run 3-D, #1 – Top-view of the Earth-fixed map and trajectory with ground truth and uncertainty bounds.

For this reason, it would be very beneficial to have a trajectory that excites equally all directions as well as larger horizontal and vertical fields of view to, respectively, better capture turns and better differentiate vertical motion. All these aspects contribute to fewer information to extract from the measurements and consequently worsen the estimation performance. Furthermore, there exists inconsistency in the attitude, which can be explained by the intrinsic nonlinearity of the problem, made more extreme by erroneous associations and movements with large rotations of the camera. Note that in the second run (Figure 18) this effect is much less noticeable. The ending result of the run is shown in Figure 15 where a top-view of the Earth-fixed map is depicted along with the estimated trajectory and ground truth, demonstrating the good performance of the overall algorithm. In this figure, it is noticeable that a number of landmarks appears inside the lab due to the presence of obstacles not indicated in the blueprint shown in Figure 16 (the lab is the first room on the left), especially the protection net (see Figure 7(e)) whose knots are identified as landmarks several times.

Figures 16, 17, and 18 depict the results of the second, larger, run. In the second run, the vehicle starts inside the usable area of the lab and, after a small lap inside the room, it is hand-driven outside the room into the corridor shown in Figure 16 travelling a total of 80 meters at an average speed of 0.4 m/s. Please note that the VICON system is only available inside the working area of the lab, and that is why ground truth disappears after around 45 s into the run. In Figure 16, a top view of the Earth-fixed map along the estimated trajectory (solid blue), shows the floor blueprint which allows a qualitative validation. In that figure, the ground truth trajectory (dashed red) obtained from the *VICON* is also shown. The trajectory starts inside the lab and ends in the middle of the corridor, marked by the blue triangle. The small ellipses are the 2-D projection of the 3σ uncertainty ellipsoids. The landmark map can be clearly related to objects in the blueprint which once more indicates the good performance of the algorithm.

The vehicle has full 3-D motion, as the attitude in Figure 18 and the vertical position in the bottom of Figure 17 reveal. In

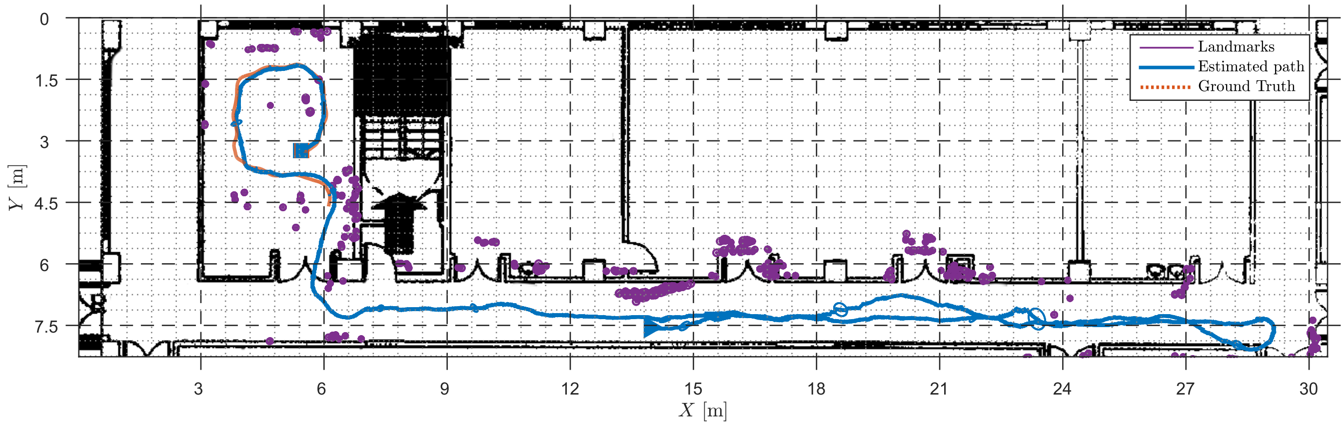


Figure 16: RB-SLAM: Run 3-D, #2 – Top-view of the Earth-fixed map and trajectory with ground truth, uncertainty bounds, and floor blueprint.

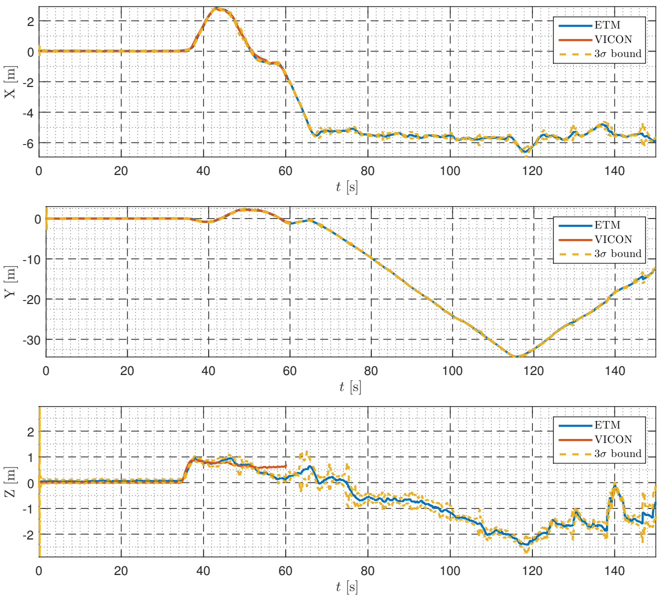


Figure 17: RB-SLAM: Run 3-D, #2 – Evolution of the position estimates with ground truth (available in the first 60 seconds) and uncertainty bounds. Horizontal trajectory (top two figures), Vertical trajectory (bottom).

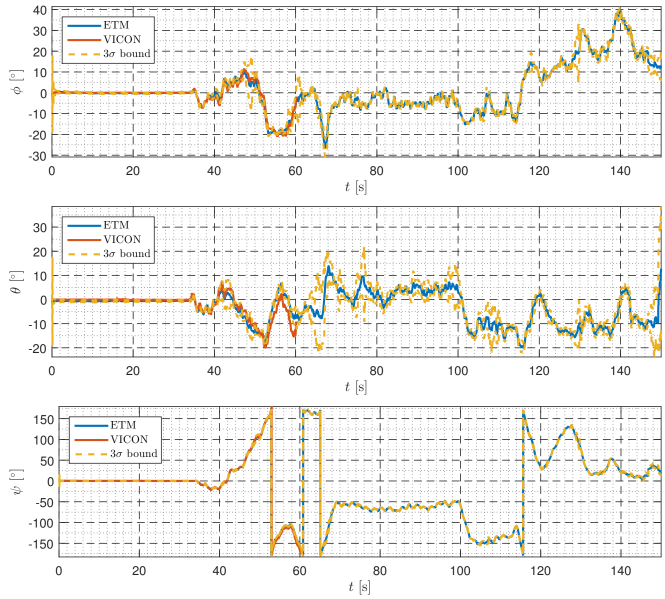


Figure 18: RB-SLAM: Run 3-D, #2 – Evolution of the Euler angles estimates with uncertainty bounds against the ground truth (available in the first 60 seconds). From top to bottom: roll, pitch, and yaw.

the latter, the Euler angles of the vehicle are depicted along with ground truth, when available, and the 3σ uncertainty bounds provided by the uncertainty description derived in this paper. As for the position estimates, once again the horizontal performance is better, as there is a drift in the vertical estimate that was not present in reality. It can be seen that the algorithm performs very well while the ground truth is available, with low and consistent uncertainty.

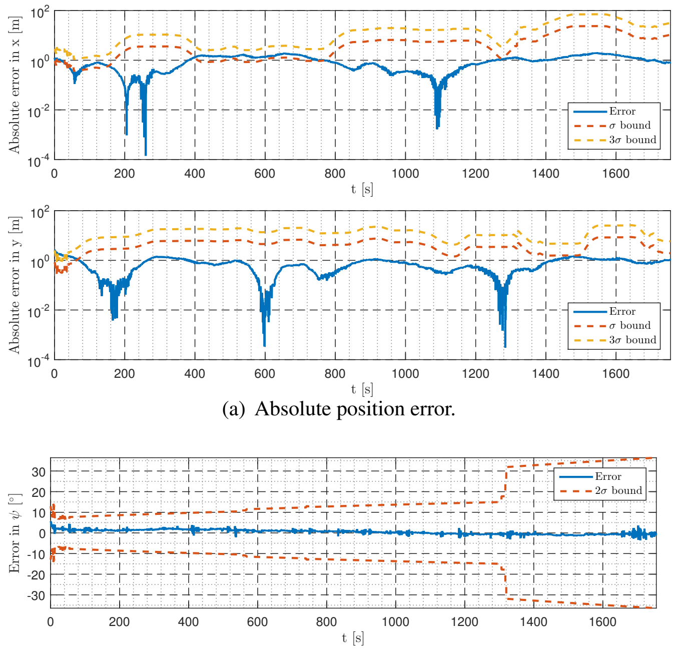
5.2. *Bearing-only SLAM*

The final part of this experimental section reports the use of the ETM algorithm in conjunction with another sensor-based SLAM filter, this time tailored for bearing-only measurements aided by linear and angular velocity measurements. The filter is proposed in [23] and is tested with real data from datasets acquired by the *Rawseeds* Project (see [26] and [27]). A black and white camera is used to survey the environment and, as in the 3-D RB-SLAM experiments above, a SURF [48] implementation allows the extraction of natural features from the obtained images. The remaining data sources are the wheel encoders which provide body-fixed linear velocity and simulated angular velocity. Even though it is possible to obtain the angular velocity (or angular displacements) from both the wheel encoders and the inertial measurement unit present in the robot, both have very high errors (standard deviations above 1.5°/s). This results in a very distorted trajectory, when dead-reckoning the body-fixed linear velocity provided by the odometry and either angular velocities. Bearing-only (and range-only) SLAM algorithms rely greatly on the quality of the ego-motion measurements that drive them, in this case the linear and angular velocities. This is particularly relevant when the field of view of the camera is limited to a region in the front of the vehicle, because, when turning, the camera will quickly lose track of features that help correct motion information. The available benchmark solution to monocular SLAM provided in the dataset, based on the algorithms in [49] and [50], shows precisely both how the monocular camera filtering results can be poor, due to the limited field-of-view and the large distance, and how the odometry is extremely deficient in turns. For these reasons, the authors decided to use an artificial measurement of angular velocity, taking the measurement provided by the scan matching procedure and adding artificial noise with a realistic standard deviation of 0.15°/s – obtained from the data in the 3-D range-and-bearing experiments detailed previously. This noise level can easily be found nowadays in off-the-shelf IMU like the *Microstrain 3DM-GX3-25*. The dataset also provides ground truth from a system based on industrial cameras, visual tags mounted on the robot, and ad hoc software in select parts of the trajectory. This ground truth is complemented by what is called extended ground truth, available for all the trajectory, which is computed using scan matching from the laser scanners in the robot. More details of the experimental setup are given in [23]. The travelled distance in this run is 774 m in 29 minutes, with an average speed of 0.4 m/s, and, since the trajectory is purely two-dimensional, only horizontal results are shown here and on the video available at [42].

Figures 19 and 20 depict, respectively, the estimation error of the pose of the vehicle and the estimated map and trajectory. In Figure 19(a), the absolute estimation error (in solid blue) is accompanied by the uncertainty bounds at σ level (dashed red) and 3σ (dashed yellow). It can be seen that the uncertainty characterization is, in some cases, very conservative, but the consistency is maintained throughout the long run. Furthermore, the error is low, with its norm averaging at 1.2 metres. The attitude in this problem is reduced to the yaw angle, whose estimation error is presented in Figure 19(b) in blue with the uncertainty bounds, clearly demonstrating consistency, even if conservative, and a low level of error (average norm of 0.4°).

Finally, in Figure 20, the estimated Earth-fixed map and the vehicle trajectory are depicted, along with the ground truth for the latter and the executive drawings of the area. To avoid overcrowding the figures with too much information, the landmark map and corresponding uncertainty ellipses are presented alone in Figure 20(a) and the trajectory, ground truth, dead-reckoned odometry, and uncertainty ellipses at fixed intervals are shown in Figure 20(b). Even though there is no ground truth for the estimated map, since the landmarks are obtained through feature detection, the executive drawings allow a good qualitative evaluation of the result. As to the trajectory, it can be observed that the complete SLAM algorithm corrects the odometry with low error, even though the last part of the map has much higher uncertainty than the rest. It should be noted that in the sensor-based BO-SLAM filter no dedicated loop closure procedures are used, which may explain this effect.

These experiments were designed to practically validate the optimization-based Earth-fixed Trajectory and Map estimation algorithm presented in this paper, as part of an integrated two-step SLAM filter, with the demonstration of consistency and good performance of the proposed uncertainty characterization. Furthermore, they allow the performance assessment of the ETM algorithm under real-world conditions, using the proposed formulation of the Procrustes problem and implementing it in each time-step to address a relevant robotics issue, SLAM.



(a) Absolute position error.

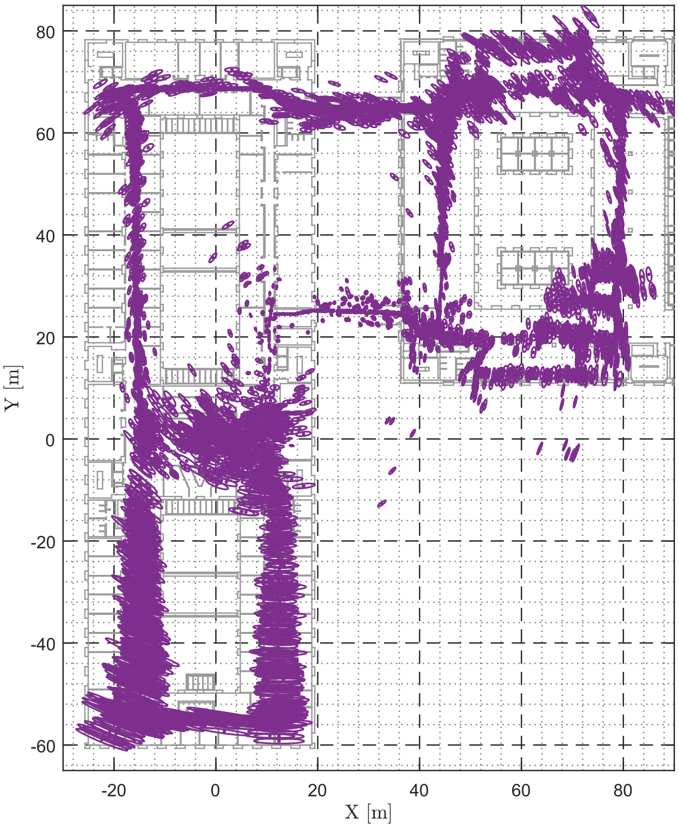
(b) Yaw error.

Figure 19: BO-SLAM: Evolution of the estimation error of the vehicle pose. On top, the two horizontal coordinates of the position estimation error, and on the bottom, the attitude estimation error in the form of the yaw error.

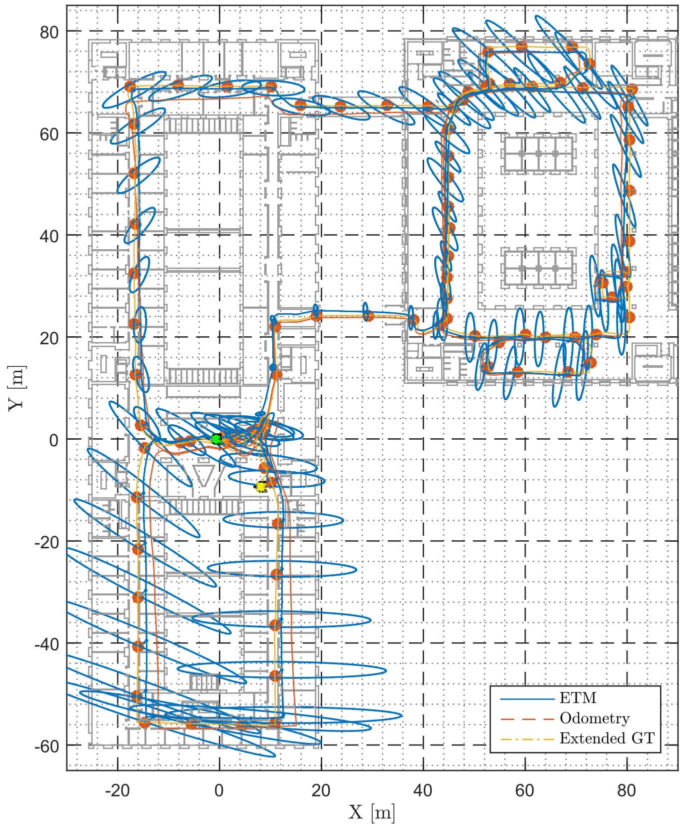
6. Conclusions

This paper proposed an optimization-based algorithm that is part of a novel methodology for simultaneous localization and mapping. The algorithm, fully characterized uncertainty-wise, provides estimates of the landmark map and of the attitude and position of the vehicle in an Earth-fixed frame, using only the body-fixed map provided by existing globally convergent filters. Building on the body-fixed map provided by the sensor-based filter, the problem of obtaining the Earth-fixed trajectory and map was formulated using an orthogonal Procrustes problem approach. The resulting optimization problem has a closed-form solution and a statistical description of the obtained Earth-fixed map is also proposed building on previous work by the authors. Furthermore, the performance and consistency of the algorithm were validated in simulation for a range-only SLAM formulation and experimentally for range-and-bearing and bearing-only formulations. These results, with ground truth data, showed also the good performance of the SLAM algorithm as a whole.

With respect to future work, the authors identify one main course of action, consisting of the optimization of the implementation for real-time, which is of paramount importance for achieving a truly online filter that can be used with autonomous



(a) Landmark map with uncertainty ellipses.



(b) Trajectory with ground truth and uncertainty.

Figure 20: BO-SLAM: The two-dimensional Earth-fixed trajectory and map, with ground truth, executive drawings, and uncertainty bounds.

ground or aerial vehicles. Furthermore, performing new range-only experiments for sensor-based SLAM to solve the technical issues found in [22] is an important future step, to have a complete experimental validation of the ETM algorithm for all SLAM formulations.

AppendixA. Error computation

Let $\mathcal{P} \in SE(3)$ represent the transformation from the body-fixed frame $\{B\}$ to the Earth-fixed frame $\{E\}$, encoding both the rotation $\mathbf{R}$ and the translation ${}^{E}\mathbf{p}$, such that

$$\begin{bmatrix} {}^{E}\mathbf{p}_i \\ 1 \end{bmatrix} = \mathcal{P} \begin{bmatrix} \mathbf{p}_i \\ 1 \end{bmatrix}.$$

Then, $\mathcal{P}$ is in fact a representation of the pose of the vehicle and is given by

$$\mathcal{P} = \begin{bmatrix} \mathbf{R} & {}^{E}\mathbf{p} \\ \mathbf{0} & 1 \end{bmatrix}.$$

It is a matter of computation to see that the inverse transformation is given by

$$\mathcal{P}^{-1} = \begin{bmatrix} \mathbf{R}^T & -\mathbf{R}^{T\,E}\mathbf{p} \\ \mathbf{0} & 1 \end{bmatrix}$$

and that

$$\mathcal{P}^{-1}\mathcal{P} = \begin{bmatrix} \mathbf{R}^T\mathbf{R} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} = \mathbf{I} = \mathcal{P}\mathcal{P}^{-1}$$

Given its properties, it makes sense to define the pose estimation error as

$${}^{B}\mathbf{E}_{\mathcal{P}} = \mathcal{P}^{-1}\hat{\mathcal{P}} = \begin{bmatrix} \mathbf{R}^T\hat{\mathbf{R}} & \mathbf{R}^T({}^{E}\hat{\mathbf{p}} - {}^{E}\mathbf{p}) \\ \mathbf{0} & 1 \end{bmatrix}.$$

In order to have the translational error expressed in the Earth-fixed frame, this error can be multiplied by the actual rotation, yielding

$$\mathbf{E}_{\mathcal{P}} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} {}^{B}\mathbf{E}_{\mathcal{P}} \begin{bmatrix} \mathbf{R}^T & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{R}}\mathbf{R}^T & {}^{E}\hat{\mathbf{p}} - {}^{E}\mathbf{p} \\ \mathbf{0} & 1 \end{bmatrix}.$$

from where the translational and rotational errors can be extracted to give

$$\mathbf{e}_p = {}^{E}\mathbf{p} - {}^{E}\hat{\mathbf{p}},$$

and

$$\mathbf{E}_R = \hat{\mathbf{R}}\mathbf{R}^T.$$

This last error function is an orthonormal matrix in SO(n), and, as such, can be parametrized by an angular error. This is a minimal representation, obtainable using the exponential map [51]

$$\mathbf{R} = \mathrm{expm}\left(\mathbf{S}(\boldsymbol{\theta})\right)$$

with $\boldsymbol{\theta} \in \mathbb{R}^{n\times(n-1)/2}$. Therefore, the rotation error is then given by

$$\mathbf{e}_\theta = \mathbf{S}^{-1}\Big(\mathrm{logm}(\hat{\mathbf{R}}\mathbf{R}^T)\Big). \tag{A.1}$$

With these definitions, it is possible to evaluate both the position and attitude estimation error of any navigation algorithm. However, since these are vector quantities, it is advantageous to obtain scalar quantities for performance evaluation. The first two are the absolute trajectory error (ATE) given by

$$ATE(k) = \|\mathbf{e}_{p_k}\| = \left\|{}^{E}\mathbf{p}_k - {}^{E}\hat{\mathbf{p}}_k\right\| \tag{A.2}$$

and the absolute attitude error (AAE) given by

$$AAE(k) = \|\mathbf{e}_{\theta_k}\| = \left\|\mathbf{S}^{-1}\Big(\operatorname{logm}(\hat{\mathbf{R}}\mathbf{R}^T)\Big)\right\|. \tag{A.3}$$

These allow an evaluation of the global consistency of the algorithm at a given point. Finally, the relative pose error measures the local accuracy of the trajectory over a fixed time interval Δ. It encodes the drift of the trajectory, something that can be useful in certain systems. This is given by

$$\begin{aligned} RPE(k,\Delta) &= \left\|\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\left(\mathcal{P}_k^{-1}\mathcal{P}_{k+\Delta}\right)^{-1}\left(\hat{\mathcal{P}}_k^{-1}\hat{\mathcal{P}}_{k+\Delta}\right)\begin{bmatrix}\mathbf{0}\\ \mathbf{I}\end{bmatrix}\right\| \\ &= \left\|\hat{\mathbf{R}}_k^T\left({}^{E}\hat{\mathbf{p}}_k - {}^{E}\hat{\mathbf{p}}_{k+\Delta}\right) - \mathbf{R}_k^T\left({}^{E}\mathbf{p}_k - {}^{E}\mathbf{p}_{k+\Delta}\right)\right\| \end{aligned} \tag{A.4}$$

which combines rotational and translational errors in a single metric, sometimes used in the literature (see, for example, [52]).

Finally, the metric for the quality of a map can be evaluated by the absolute landmark estimation error

$$ALE(i,k,F) = \|{}^{F}\mathbf{p}_i - {}^{F}\hat{\mathbf{p}}_i\|,$$

for a particular landmark i expressed in frame $\{F\}$. When evaluating the complete map, the absolute map error is

$$AME(k,F) = \frac{1}{N}\sum_{i=1}^{N} ALE(i,k,F).$$

Naturally, each of these metrics can then be averaged through the run, its standard deviation and root mean square also computed to provide a better and comparable overall view of the performance.

Acknowledgments

This work was supported by the Fundação para a Ciência e Tecnologia (FCT) through LARSyS - FCT Pluriannual funding 2020-2023, through IDMEC, under LAETA UID/EMS/50022/2013, through CTS (UID/EEA/00066/2019) contracts, and through project REPLACE (PTDC/EEI-AUT/32107/2017), by the University of Macau Project MYRG2018-00198-FST, and by the Macao Science and Technology Development Fund under Grant FDCT/026/2017/A1. The work of P. Lourenço was partially supported by the FCT PhD Student Grant SFRH/BD/89337/2012.

C. Silvestre is with the Department of Electrical and Computer Engineering of the Faculty of Science and Technology of the University of Macau, Macau, China, on leave from Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal.

References

[1] J. Qi, D. Song, H. Shang, N. Wang, C. Hua, C. Wu, X. Qi, J. Han, Search and rescue rotary-wing uav and its application to the lushan ms 7.0 earthquake, Journal of Field Robotics 33 (3) (2016) 290–321. doi:10.1002/rob.21615.

[2] D. Kingston, R. W. Beard, R. S. Holt, Decentralized Perimeter Surveillance Using a Team of UAVs, IEEE Transactions on Robotics 24 (6) (2008) 1394–1404. doi:10.1109/TRO.2008.2007935.

[3] G. Christie, A. Shoemaker, K. Kochersberger, P. Tokekar, L. McLean, A. Leonessa, Radiation search operations using scene understanding with autonomous UAV and UGV, Journal of Field Roboticsdoi:10.1002/rob.21723.

[4] L. Merino, F. Caballero, J. Martínez-de Dios, J. Ferruz, A. Ollero, A cooperative perception system for multiple UAVs: Application to automatic detection of forest fires, Journal of Field Robotics 23 (3-4) (2006) 165–184. doi:10.1002/rob.20108.

[5] J. Katrasnik, F. Pernus, B. Likar, A Survey of Mobile Robots for Distribution Power Line Inspection, IEEE Transactions on Power Delivery 25 (1) (2010) 485–493. doi:10.1109/TPWRD.2009.2035427.

[6] F. Kendoul, Survey of advances in guidance, navigation, and control of unmanned rotorcraft systems, Journal of Field Robotics 29 (2) (2012) 315–378. doi:10.1002/rob.20414.

[7] H. Durrant-Whyte, T. Bailey, Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms, IEEE Robotics & Automation Magazine 13 (2) (2006) 99–110. doi:10.1109/MRA.2006.1638022.

[8] T. Bailey, H. Durrant-Whyte, Simultaneous localization and mapping (SLAM): Part II, IEEE Robotics & Automation Magazine 13 (3) (2006) 108–117. doi:10.1109/MRA.2006.1678144.

[9] J. Fuentes-Pacheco, J. Ruiz-Ascencio, J. M. Rendón-Mancha, Visual simultaneous localization and mapping: a survey, Artificial Intelligence Review 43 (1) (2015) 55–81. doi:10.1007/s10462-012-9365-8.

[10] S. Huang, G. Dissanayake, A critique of current developments in simultaneous localization and mapping, International Journal of Advanced Robotic Systems 13 (5) (2016) 1–13. doi:10.1177/1729881416669482.

[11] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, J. J. Leonard, Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age, IEEE Transactions on Robotics 32 (6) (2016) 1309–1332. doi:10.1109/TRO.2016.2624754.

[12] T. Bailey, J. Nieto, J. Guivant, M. Stevens, E. Nebot, Consistency of the EKF-SLAM Algorithm, in: 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 3562–3568.

[13] S. Julier, J. Uhlmann, A counter example to the theory of simultaneous localization and map building, in: Proc. of the 2001 IEEE Int. Conf. on Robotics and Automation (ICRA), Vol. 4, Seul, South Korea, 2001, pp. 4238–4243.

[14] S. Huang, G. Dissanayake, Convergence and Consistency Analysis for Extended Kalman Filter Based SLAM, IEEE Transactions on Robotics 23 (5) (2007) 1036–1049.

[15] A. Bishop, P. Jensfelt, A Stochastically Stable Solution to the Problem of Robocentric Mapping, in: Proc. of the IEEE International Conference on Robotics and Automation, Kobe, Japan, 2009, pp. 1615–1622.

[16] E. Jones, S. Soatto, Visual-inertial navigation, mapping and localization: A scalable real-time causal approach., The International Journal of Robotics Research 30 (4) (2011) 407–430.

[17] J. Kelly, G. S. Sukhatme, Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration, The International Journal of Robotics Research 30 (1) (2011) 56–79.

[18] J. Castellanos, R. Martinez-Cantin, J. Tardós, J. Neira, Robocentric map joining: Improving the consistency of EKF-SLAM, Robotics and Autonomous Systems 55 (1) (2007) 21–29. doi:10.1016/j.robot.2006.06.005.

[19] J. Yang, A. Dani, S.-J. Chung, S. Hutchinson, Vision-based localization and robot-centric mapping in riverine environments, Journal of Field Robotics 34 (3) (2017) 429–450. doi:10.1002/rob.21606.

[20] B. J. Guerreiro, P. Batista, C. Silvestre, P. Oliveira, Globally Asymptotically Stable Sensor-based Simultaneous Localization and Mapping, IEEE Transactions on Robotics 29 (6) (2013) 1380–1395. doi:10.1109/TRO.2013.2273838.

[21] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, Simultaneous Localization and Mapping for Aerial Vehicles: a 3-D sensor-based GAS filter, Autonomous Robots 40 (2016) 881–902. doi:10.1007/s10514-015-9499-z.

[22] P. Lourenço, P. Batista, P. Oliveira, C. Silvestre, C. L. P. Chen, Sensor-based Globally Exponentially Stable Range-Only Simultaneous Localization and Mapping, Robotics and Autonomous Systems 68 (2015) 72–85. doi:10.1016/j.robot.2015.01.010.

[23] P. Lourenço, P. Batista, P. Oliveira, C. Silvestre, A Globally Exponentially Stable filter for Bearing-Only Simultaneous Localization and Mapping with Monocular vision, Robotics and Autonomous Systems 100 (2018) 61–77. doi:10.1016/j.robot.2017.11.001.

[24] B. J. Guerreiro, P. Batista, C. Silvestre, P. Oliveira, Sensor-based Simultaneous Localization and Mapping - Part II: Online Inertial Map and Trajectory Estimation, in: Proceedings of the 2012 American Control Conference, Montréal, Canada, 2012, pp. 6334–6339.

[25] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, 3-D Inertial Trajectory and Map Online Estimation: Building on a GAS Sensor-based SLAM filter, in: Proc. of the 2013 European Control Conference, Zurich, Switzerland, 2013, pp. 4214–4219. doi:10.23919/ECC.2013.6669679.

[26] A. Bonarini, W. Burgard, G. Fontana, M. Matteucci, D. G. Sorrenti, J. D. Tardos, Rawseeds: Robotics advancement through web-publishing of sensorial and elaborated extensive data sets, in: Proceedings of IROS'06 Workshop on Benchmarks in Robotics Research, 2006.

[27] S. Ceriani, G. Fontana, A. Giusti, D. Marzorati, M. Matteucci, D. Migliore, D. Rizzi, D. G. Sorrenti, P. Taddei, Rawseeds ground truth collection systems for indoor self-localization and mapping, Autonomous Robots 27 (4) (2009) 353–371. doi:10.1007/s10514-009-9156-5.

[28] C. Goodall, Procrustes Methods in the Statistical Analysis of Shape, Journal of the Royal Statistical Society. Series B (Methodological) 53 (2) (1991) 285–339.

[29] S. Umeyama, Least-squares estimation of transformation parameters between two point patterns, IEEE Transactions On Pattern Analysis and Machine Intelligence 13 (4) (1991) 376–380.

[30] B. K. P. Horn, H. Hilden, S. Negahdaripour, Closed-form solution of absolute orientation using orthonormal matrices, Journal of the Optical Society America 5 (7) (1988) 1127–1135.

[31] R. Sibson, Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling, Journal of the Royal Statistical Society. Series B (Methodological) 41 (2) (1979) 217–229.

[32] R. Sibson, Studies in the robustness of multidimensional scaling: Procrustes statistics, Journal of the Royal Statistical Society. Series B (Methodological) 40 (2) (1978) 234–238.

[33] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, Uncertainty Characterization of the Orthogonal Procrustes Problem with Arbitrary Covariance Matrices, Pattern Recognition 61 (2017) 210–220. doi:10.1016/j.patcog.2016.07.037.

[34] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, New Design Techniques for Globally Convergent Simultaneous Localization and Mapping: Analysis and Implementation, in: T. I. Fossen, K. Y. Pettersen, H. Nijmeijer (Eds.), Sensing and Control for Autonomous Vehicles: Applications to Land, Water and Air Vehicles, Lecture Notes in Control and Information Sciences, Springer, 2017. doi:10.1007/978-3-319-55372-6_6.

[35] R. E. Kalman, A new approach to linear filtering and prediction problems, Transactions of the ASME–Journal of Basic Engineering 82 (Series D) (1960) 35–45.

[36] A. Gelb, Applied Optimal Estimation, MIT Press, 1974.

[37] J. C. Gower, G. B. Dijksterhuis, Procrustes Problems, Oxford Statistical Science Series, Oxford University Press, 2004.

[38] P. Schönemann, R. Carroll, Fitting One Matrix to Another Under Choice of a Central Dilation and a Rigid Motion, Psychometrika 35 (1970) 245–255.

[39] J. L. Crassidis, F. L. Markley, Y. Cheng, Survey of Nonlinear Attitude Estimation Methods, Journal of Guidance, Control, and Dynamics 30 (1) (2007) 12–28. doi:10.2514/1.22452.

[40] T. A. Johansen, E. Brekke, Globally Exponentially Stable Kalman Filtering for SLAM with AHRS, in: Proceedings of the 19th International Conference on Information Fusion (FUSION 2016), Heidelberg. Germany, 2016, pp. 909–916.

[41] P. Lourenço, Globally Convergent Simultaneous Localization and Mapping: Design Techniques, Analysis, and Implementation, PhD Dissertation, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal (Jan. 2019).
URL http://web.tecnico.ulisboa.pt/~ist162690/docs/thesis/2018_PhD_PL.pdf

[42] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, Earth-fixed Trajectory and Map Online Estimation: Building on GES sensor-based SLAM filters, https://youtu.be/d2dtR56eA9g.

[43] T. Pavlidis, S. L. Horowitz, Segmentation of plane curves, IEEE Transactions on Computers C-23 (8) (1974) 860–870. doi:10.1109/T-C.1974.224041.

[44] C. Fernández, V. Moreno, B. Curto, J. A. Vicente, Clustering and line detection in laser range measurements, Robotics and Autonomous Systems 58 (5) (2010) 720–726. doi:10.1016/j.robot.2009.10.008.

[45] J. Neira, J. Tardós, Data Association in Stochastic Mapping Using the Joint Compatibility Test, IEEE Transactions on Robotics and Automation 17 (6) (2001) 890–897.

[46] VICON Motion Systems, Essentials of motion capture v1.2, http://www.udel.edu/PT/Research/MAL/essentials_of_motion_capture_v1_2.pdf.

[47] M. Windolf, N. Götzen, M. Morlock, Systematic accuracy and precision analysis of video motion capturing systems - exemplified on the vicon-460 system, Journal of Biomechanics 41 (12) (2008) 2776–2780.

[48] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, Speeded-Up Robust Features (SURF), Computer Vision and Image Understanding 110 (3) (2008) 346–359.

[49] J. Civera, A. J. Davison, J. M. M. Montiel, Inverse depth parametrization for monocular slam, IEEE Transactions on Robotics 24 (5) (2008) 932–945. doi:10.1109/TRO.2008.2003276.

[50] J. Civera, O. G. Grasa, A. J. Davison, J. M. M. Montiel, 1-Point RANSAC for extended Kalman filtering: Application to real-time structure from motion and visual odometry, Journal of Field Robotics 27 (5) (2010) 609–631. doi:10.1002/rob.20345.

[51] J. Craig, Introduction to Robotics: Mechanics and Control, Addison-Wesley series in electrical and computer engineering: control engineering, Pearson/Prentice Hall, 2005.

[52] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers, A benchmark for the evaluation of rgb-d slam systems, in: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 573–580. doi:10.1109/IROS.2012.6385773.