ROBOT LOCALIZATION AND 3D MAPPING: OBSERVABILITY ANALYSIS AND APPLICATIONS

Luca Carlone[1], **Vito Macchia**[2], **Federico Tibaldi**[1], **and Basilio Bona**[1]

[1]*Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy,*
{luca.carlone, federico.tibaldi, basilio.bona}@polito.it
[2]*Istituto Superiore Mario Boella, Via Pier Carlo Boggio, 61, 10138 Torino, Italy,*
vito.macchia@ismb.it

ABSTRACT

In this work we investigate a quaternion-based formulation of 3D Simultaneous Localization and Mapping with Extended Kalman Filter (EKF-SLAM) using relative pose measurements. The equations of the filter do not rely on heuristic solutions for preserving the unit norm of quaternions of rotation, nor introduce spurious measurements to force the unitary norm constraint. The proposed model is formally shown to be *completely observable* using standard tools for observability analysis in piecewise linear systems. Moreover, we report numerical results of the application of the proposed approach on real data from the *Rawseeds dataset*. The contribution is motivated by the possibility of abstracting multi-sensorial information in terms of relative pose measurements and for its straightforward extensions to the multi robot case.

Key words: 3D EKF-SLAM, Unit quaternions, Observability, Relative pose measurements.

1. INTRODUCTION

In several application scenarios, ranging from exploration to search and rescue, the construction of a world model and the concurrent estimation of robot location is crucial for mission accomplishment, for enhancing motion planning effectiveness and for attaining a desired level of situational awareness. This estimation problem is usually referred to as *Simultaneous Localization and Mapping* (SLAM). In the last decade, the diffusion of autonomous robots acting on uneven terrains and the use of Unmanned Autonomous Vehicles (UAV) have stressed the importance of extending the SLAM landscape to three-dimensional scenarios. In classic EKF-SLAM implementations the state vector comprises robot pose at current time and the *position* of *landmarks* in the environment. Recent examples of this strategy are the articles [1, 2]. The complexity of the filter is quadratic in the number of *landmarks*, hence a naive implementation prevents large scale mapping due to the overwhelming computational cost [3]. It is worth noticing that including the position of each landmark in the state vectors requires to select few significative elements of the environment to be modeled, whereas it may be desirable to have a dense representation of the environment, which can be a more reliable basis for path planning and model-based navigation. Another interpretation of EKF-SLAM leads to include in the state vector (instead of the point features) the vantage poses of the robot (*observation poses*), from which such features are detected. This allows making the state dimension independent on the number of features observed from a single observation pose. The corresponding filter is usually referred to as *delayed-state* EKF-SLAM [4], *view-based* EKF [5] or *pose-based* SLAM [6].

In this article we report the results of our research on pose-based 3D SLAM with Extended Kalman filter. The contribution of the article is threefold. First of all, we report a quaternion-based formulation of EKF-SLAM; both prediction and update phase of our EKF are based on relative pose measurements. The equations do not rely on heuristic solutions for preserving the unit norm of quaternions of rotation, nor use the *projection filter* [7], which introduces spurious measurements to force the unitary norm constraint. The effectiveness of the quaternions of rotation lies in their robustness (i.e., *singularity-free*), since no three-parameters representation for rotation (e.g., roll-pitch-yaw or Euler angles [4]) can be both global and non-singular. As a second contribution we present the results of our observability study of the proposed SLAM model. We show that the system is not *locally observable*, although it is *completely observable* over a given interval of time, as long as this interval includes an observation of the *anchor*, i.e., the landmark that was set to the origin of the reference frame for mapping. The third contribution is the testing of the proposed approach on real data.

2. PROBLEM STATEMENT AND MOTIVATIONS

An agent (mobile robot, autonomous vehicle, UAV, etc.) moves in an unstructured non-planar environment with the primary aim of building a consistent representation of the scenario while estimating its own position. The agent is equipped with proprioceptive and exteroceptive sensors. The formers allow to acquire information on the self-motion of the robot (*odometry*); the latter provides measurements of the surrounding scenario. The odometric information may be provided, for instance, by wheel encoders in planar scenarios, by scan matching procedures, by Inertial Measurement Units, or by vision sensors. All previous sensor data can be easily abstracted in terms of measurement of the relative pose change (roto-translation) between two consecutive sensor measurements. The sensors that are typically used for exteroceptive sensing in 3D robotic navigation are more commonly stereo, monocular, depth cameras, or 3D laser scanners. Such devices produce a large amount of data that are usually processed by an *information synthesis* (preprocessing) block before being included in the SLAM estimation process. The outcome of information synthesis, in common applications, can be roughly classified according to the following categories: (i) *position* of 3D points corresponding to distinguishable point features, (ii) *orientation* of objects (e.g., normal to planes, direction of edges), (iii) *bearing* to landmarks (e.g., azimuth and elevation at which a far landmark is observed), (iv) *relative poses* with respect to 3D objects (e.g., other vehicles, distinguishable rigid bodies).

We now want to point out that all the previous categories can be easily abstracted in terms of relative pose measurements. For the last category in the list this abstraction is straightforward. Regarding the first three categories we can notice that (i) positions of points, (ii) objects orientations and (iii) bearings to far landmarks can be modeled by means of 3D vectors. In particular line directions, normal to planes, and other examples falling in the second and third categories may be represented as *unit vectors* (since they provide only a relative orientation information), whereas positions are described by 3D vectors in the robot frame. Therefore, we can associate, to each observation pose, a collection of 3D vectors that provides a compact representation of robot perception. Moreover, *registration techniques* (see [8]) can be successfully employed for retrieving the relative transformation (roto-translation) between two sets of 3D vectors, hence to estimate the relative pose between two observation poses. Therefore, all the listed exteroceptive measurements can be treated in terms of relative pose measurements. This motivates our investigation of a 3D SLAM approach based on relative pose measurements. We remark that 3D vectors are also an effective model for other exteroceptive perceptions such as star-trackers, sun sensors, or measurements of the gravity vector, enforcing the generality of a relative pose-based SLAM framework.

3. EKF-SLAM FROM RELATIVE POSE MEASUREMENTS

We consider a situation in which the state vector at time k includes the robot pose $x_0(k) \doteq [p_0(k)^\top \; q_0(k)^\top]^\top$ and the pose of n *landmarks* $x_j(k) \doteq [p_j(k)^\top \; q_j(k)^\top]^\top, \; j = 1, \ldots, n$ (past poses assumed by the robot). In particular, the generic pose $i = 0, 1, \ldots, n$ comprises a Cartesian position $p_i(k) \in \mathbb{R}^4$ (in homogeneous coordinates) and a unit quaternion $q_i(k)$. Hence the overall state vector for SLAM is $x(k) \doteq [x_0^\top(k) \ldots x_n^\top(k)]^\top$. Sometimes we indicate with Ω the set including both robot and landmark poses ($|\Omega| = n + 1$). As usual in Kalman Filter literature, the index $(k|k-1)$ labels quantities that incorporate information of the process model (e.g., odometry), whereas the label $(k|k)$ marks quantities that include also the measurements acquired at time k. Therefore, the estimate of the state after the prediction phase (*prior*) is $\hat{x}(k|k-1)$, and the corresponding covariance matrix is $P(k|k-1)$; the estimate of the state after the update phase (*posterior*) is instead $\hat{x}(k|k)$, and the corresponding covariance matrix is $P(k|k)$. The *estimation error* for a generic pose $i \in \Omega$ is further defined as $\tilde{x}_i(k|k) \doteq [\tilde{t}_i(k|k)^\top \; \tilde{\theta}_i(k|k)^\top]^\top \in \mathbb{R}^6$, where $\tilde{t}_i(k|k)^\top$ is the Cartesian error, i.e., $\tilde{p}_i(k|k) = [\tilde{t}_i(k|k)^\top \; 1]^\top$, where $\tilde{p}_i(k|k) = p_i(k) \ominus \hat{p}_i(k|k)$ (homogeneous difference between actual and estimated position); moreover, given the orientation error expressed (in quaternion form) as $\tilde{q}_i(k|k) = \hat{q}_i(k|k)^{-1} q_i(k)$, the orientation error vector $\tilde{\theta}_i(k|k)^\top$ is such that $\tilde{q}_i(k|k) \approx [\tilde{\theta}_i(k|k)^\top/2 \;\; 1]^\top$. Accordingly, the *error state vector* is $\tilde{x}(k|k) \doteq [\tilde{x}_0(k|k)^\top \ldots \tilde{x}_n(k|k)^\top]^\top \in \mathbb{R}^{6(n+1)}$ and the corresponding covariance is $P(k|k) \in \mathbb{R}^{6(n+1)\times 6(n+1)}$.

3.1. Prediction Phase

The prediction phase is aimed at propagating the belief on robot pose according to the odometric information $u(k) = [p_u(k)^\top q_u(k)^\top]^\top$, being $p_u(k)$ and $q_u(k)$ the translation vector and the quaternion describing the pose change between time $k-1$ and time k, respectively. It is worth noticing that the information $u(k)$ is expressed in the local frame of the robot at time $k-1$. As usual in EKF framework, we assume to have a probabilistic description of the odometry information, i.e., the measured value, namely $\bar{u}(k)$, and the covariance matrix of measurement noise, $P_u(k) \in \mathbb{R}^{6\times 6}$. A general process model for robot motion, which describes the current robot pose $x_0(k)$, given the previous pose $x_0(k-1)$ and the odometric information $u(k)$, can be written as follows:

$$\begin{cases} p_0(k) = p_0(k-1) \oplus \left(q_0(k-1)p_u(k)q_0(k-1)^{-1}\right) \\ q_0(k) = q_0(k-1)q_u(k) \end{cases} \tag{1}$$

where $\oplus$ is the sum in homogeneous coordinates. Therefore, the *prior* for the pose of the robot is:

$$\begin{cases} \hat{p}_0(k|k-1) = \hat{p}_0(k-1|k-1) \oplus (\hat{q}_0(k-1|k-1) \\ \qquad\qquad \bar{p}_u(k)\hat{q}_0(k-1|k-1)^{-1}\right) \\ \hat{q}_0(k|k-1) = \hat{q}_0(k-1|k-1)\bar{q}_u(k) \end{cases} \tag{2}$$

Assuming that the pose of the landmarks does not change (*static environment*), for a generic landmark j the prediction phase is simply:

$$\begin{cases} \hat{p}_j(k|k-1) = \hat{p}_j(k-1|k-1) \\ \hat{q}_j(k|k-1) = \hat{q}_j(k-1|k-1) \end{cases}$$

In order to devise the covariance matrix after prediction we need to linearize the process model, obtaining the *error state process model* [9], in which the prediction phase is expressed in terms of estimation errors. We omit the complete derivation for space reasons, and we only report the final result. The *error state process model* can be written as:

$$\tilde{x}(k|k-1) = \mathcal{F}_k\tilde{x}(k-1|k-1) + \mathcal{G}_k\tilde{u}(k) \tag{3}$$

with:

$$\mathcal{F}_k = \begin{bmatrix} F_k & O_{6,6} & \dots & O_{6,6} \\ O_{6,6} & I_6 & \ddots & \vdots \\ \vdots & \ddots & \ddots & O_{6,6} \\ O_{6,6} & \dots & O_{6,6} & I_6 \end{bmatrix} \quad \mathcal{G}_k = \begin{bmatrix} G_k \\ O_{6,6} \\ \vdots \\ O_{6,6} \end{bmatrix} \tag{4}$$

$$F_k = \begin{bmatrix} I_3 & -\hat{R}_{k-1}S_{u,k} \\ 0_{3,3} & R_{u,k}^\top \end{bmatrix}, \qquad G_k = \begin{bmatrix} \hat{R}_{k-1} & 0_{3,3} \\ 0_{3,3} & I_3 \end{bmatrix}$$

with $\hat{R}_{k-1} = R(\hat{q}_0(k-1|k-1))$, $R_{u,k} = R\left(\bar{q}_u(k)\right)$, and $S_{u,k} = S(\bar{p}_u(k))$, where the function $R(q)$ returns a rotation matrix corresponding to the quaternion q, while the function $S(u)$ returns a 3×3 skew-symmetric matrix built from the vector $u \in \mathbb{R}^3$.
From equation (3) it is now easy to write a first-order expression for the covariance of the estimation error of the overall SLAM state after prediction:

$$P(k|k-1) = \mathcal{F}_k P(k-1|k-1)\mathcal{F}_k^\top + \mathcal{G}_k P_u(k)\mathcal{G}_k^\top$$

3.2. Update Phase

The update phase allows including measurements information in the prior $\hat{x}(k|k-1)$. In this context we are interested in the case in which data from exteroceptive sensors can be abstracted in the form of relative pose measurements. Let us define the set of available measurements $\mathcal{M}$, such that, if $j \in \mathcal{M}$, then the robot measured the relative pose with respect to the j-th landmark, namely $z_j(k) \doteq [p_{z_j}(k)^\top \;\; q_{z_j}(k)^\top]^\top$; then, the j-th measurement can be written as:

$$\begin{cases} \bar{p}_{z_j}(k) \oplus \tilde{p}_{z_j}(k) = q_0(k)^{-1}\left(p_j(k) \ominus p_0(k)\right) q_0(k) \\ \bar{q}_{z_j}(k)\tilde{q}_{z_j}(k) = q_0(k)^{-1}q_j(k) \end{cases} \tag{5}$$

where $\bar{z_j}(k) = [\bar{p}_{z_j}(k)^\top \;\; \bar{q}_{z_j}(k)^\top]^\top$ is the actual measurement, $\tilde{p}_{z_j}(k) = [\tilde{t}_{z_j}(k)^\top \quad 1]^\top$ and $\tilde{q}_{z_j}(k) \approx [\tilde{\theta}_{z_j}(k)^\top/2 \quad 1]^\top$ are random variables describing measurement error. As in the prediction phase, a probabilistic description of the measurement error $\tilde{z}_j = [\tilde{t}_{z_j}(k)^\top \quad \tilde{\theta}_{z_j}(k)^\top]^\top$ is supposed to be available, in the form of the covariance matrix $P_{z_j}(k) \in \mathbb{R}^{6\times6}$, $j \in \mathcal{M}$.

To devise the equations for the EKF-SLAM update we need to compute a first-order expression for the measurement model, in terms of the involved errors. First of all, we need to compute the *innovation for the j-th measurement*, that represents the mismatch between the expected and the actual measurement. A first-order approximation of the *innovation for the j-th measurement* $\tilde{r}_j$ can be written as:

$$\tilde{r}_j(k) \approx H_j\tilde{x} - \tilde{z}_j \tag{6}$$

where:

$$H_j = \begin{bmatrix} \begin{bmatrix} -\hat{R}_k^\top & \hat{S}_k \\ 0_{3,3} & -\hat{R}_{\Delta_k}^\top \end{bmatrix} \cdots 0_{6,6} \cdots \begin{bmatrix} \hat{R}_k^\top & O_{3,3} \\ 0_{3,3} & I_3 \end{bmatrix} \cdots 0_{6,6} \cdots \end{bmatrix}$$

$$\tilde{x} = \begin{bmatrix} \tilde{t}_0(k|k-1) \\ \tilde{\theta}_0(k|k-1) \\ \vdots \\ \tilde{t}_j(k|k-1) \\ \tilde{\theta}_j(k|k-1) \\ \vdots \end{bmatrix}, \qquad \tilde{z}_j = \begin{bmatrix} \tilde{t}_{z_j}(k) \\ \tilde{\theta}_{z_j}(k) \end{bmatrix} \tag{7}$$

$$\begin{aligned} \hat{R}_k &= R\left(\hat{q}_0(k|k-1)\right) \\ \hat{R}_{\Delta_k} &= R\left(\hat{q}_0(k|k-1)^{-1}\hat{q_j}(k|k-1)\right) \\ \hat{S}_k &= S\left(R^\top\left(\hat{q}_0(k|k-1)\right)\left(\hat{t}_j(k|k-1) - \hat{t}_0(k|k-1)\right)\right) \end{aligned} \tag{8}$$

Stacking the matrices H_j in a single matrix $\mathcal{H}_k$ we obtain an expression for the overall *innovation* in function of the estimation and measurement errors:

$$\tilde{r}(k) \approx \mathcal{H}_k \tilde{x} - \tilde{z} \tag{9}$$

From equation (9), it is now possible to derive *the covariance matrix for the innovation*:

$$P_r(k) = \mathcal{H}_k P(k|k-1) \mathcal{H}_k^\top + P_z(k)$$

with $P_z(k)$ being the covariance matrix of all the available measurements $z(k)$, i.e., $P_z(k) = \text{diag}(P_{z_1}(k), P_{z_2}(k), \ldots, P_{z_{|\mathcal{M}|}}(k))$. Hence, the *Kalman gain* can be computed as [10]:

$$\mathcal{K}_k = P(k|k-1) \mathcal{H}_k^\top P_r(k)^{-1} \tag{10}$$

Given the innovation $\tilde{r}(k)$ and the gain (10) the correction term for the state can be written as $c(k) = \mathcal{K}_k \tilde{r}(k)$. Now, if we separate the Cartesian and the orientation variables in $c(k)$ for the generic pose $i \in \Omega$, namely $\tilde{t}_{c_i}(k)$ and $\tilde{\theta}_{c_i}(k)$ and we obtain the homogeneous coordinates corresponding to $\tilde{t}_{c_i}(k)$, say $p_{c_i}(k)$, and the quaternion corresponding to $\tilde{\theta}_{c_i}(k)$, say $q_{c_i}(k)$, the update equations for the state vector can be written as:

$$\begin{cases} \hat{p}_i(k|k) = \hat{p}_i(k|k-1) \oplus p_{c_i}(k) \\ \hat{q}_i(k|k) = \hat{q}_i(k|k-1) q_{c_i}(k) \end{cases} \quad i = 0, \ldots, n$$

Moreover the covariance matrix for the posterior is:

$$P(k|k) = \left(I_{6(n+1)} - \mathcal{K}_k \mathcal{H}_k\right) P(k|k-1)$$

In the following section we discuss the observability properties of the proposed EKF-based approach.

4. OBSERVABILITY ANALYSIS

Observability analysis and consistency verification for the standard EKF-SLAM formulation in planar scenarios can be found in [11, 12]. Observability analysis in 3D problems has been mainly studied in the aerospace literature [13]. In [14] the authors studied observability in 3D SLAM, using Euler angles. The work [15] reports an observability analysis of 3D SLAM based on IMU sensors within an adaptive Kalman filter.

We now present an observability analysis for the proposed model. The system under analysis is clearly nonlinear and, in this context, we employ the machinery introduced in [13] (and applied in a planar SLAM problem in [12]) for evaluating the observability properties of the piecewise linear model used in EKF-SLAM. According to the derivation of Section 3.1 and 3.2 the system can be modeled through the following time-variant linear system:

$$\begin{cases} \tilde{x}(k) = \mathcal{F}_k \tilde{x}(k-1) + \mathcal{G}_k \tilde{u}(k) \\ \tilde{r}(k) = \mathcal{H}_k \tilde{x}(k) - \tilde{z}(k) \end{cases} \tag{11}$$

The subscript k of systems matrixes $(\mathcal{F}_k, \mathcal{G}_k, \mathcal{H}_k)$ remarks their dependence on time, and for each discrete-time segment, the system can be approximated by the previous linear model. Without loss of generality let us assume that, at each pose, the robot initializes a new 3D landmark with the current observation pose; moreover the robot, at time k, is able to measure the relative pose with respect to the landmark initialized at time $k-1$. According to the results reported in [2, 13] we can study the local observability properties of the system with linear algebra tools. In particular the system is said to be *locally (instantaneously) observable* at time k if the matrix

$$\mathcal{O}_k = \left[\; (\mathcal{H}_k)^\top \; (\mathcal{H}_k \mathcal{F}_k)^\top \; (\mathcal{H}_k \mathcal{F}_k^2)^\top \; \ldots \; (\mathcal{H}_k \mathcal{F}_k^{\eta-1})^\top \; \right]^\top \tag{12}$$

is full rank (i.e., the rank is equal to the dimension of the state space, namely η). The matrix $\mathcal{O}_k$ is usually referred to as *local observability matrix* (LOM). Roughly speaking, the concept of local observability is connected to the possibility of instantaneously distinguishing the state $\tilde{x}_k$ (absolute poses of robot and landmarks in our application) using the current observation. For the study of local observability let us write the LOM in explicit form:

$$\mathcal{O}_k = \begin{bmatrix} \begin{bmatrix} -\hat{R}_k^\top & \hat{S}_k \\ O_{3,3} & -\hat{R}_{\Delta_k}^\top \end{bmatrix} & \begin{bmatrix} \hat{R}_k^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} & O_{6,6} \ldots \\ \begin{bmatrix} -\hat{R}_k^\top & M_{k,k-1}^1 \\ O_{3,3} & N_{k,k-1}^1 \end{bmatrix} & \begin{bmatrix} \hat{R}_k^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} & O_{6,6} \ldots \\ \vdots & \vdots & \vdots \\ \begin{bmatrix} -\hat{R}_k^\top & M_{k,k-1}^{\eta-1} \\ O_{3,3} & N_{k,k-1}^{\eta-1} \end{bmatrix} & \begin{bmatrix} \hat{R}_k^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} & O_{6,6} \ldots \end{bmatrix}$$

where $M_{k,k-1}^i = \hat{R}_k^\top \hat{R}_{k-1} S_{u,k} \sum_{j=0}^{i-1} \left(R_{u,k}^\top\right)^j + \hat{S}_k \left(R_{u,k}^\top\right)^i$ and $N_{k,k-1}^i = -\hat{R}_{\Delta_k}^\top \left(R_{u,k}^\top\right)^i$. In this case the (column) rank of the LOM can be easily seen to be

less than η, since the last columns of the matrix contain only zeros. This result confirms the fact that the absolute pose of the robot is not observable using only the current observation w.r.t. the previous pose. We can now introduce a milder definition of observability that allows taking into consideration all the measurements collected in a given interval of time, say $[1, k]$: in this case we are interested in studying if all the observations acquired by the robot since the initial time are sufficient for the estimation of the absolute poses included in the state vector. For studying the observability over the interval $[1, k]$, we introduce the *total observability matrix* (TOM):

$$\mathcal{O}_{[1:k]} = \begin{bmatrix} \mathcal{O}_1 \\ \mathcal{O}_2\mathcal{F}_1^{\eta-1} \\ \mathcal{O}_3\mathcal{F}_2^{\eta-1}\mathcal{F}_1^{\eta-1} \\ \vdots \\ \mathcal{O}_k\mathcal{F}_{k-1}^{\eta-1}\mathcal{F}_{k-2}^{\eta-1}\dots\mathcal{F}_1^{\eta-1} \end{bmatrix} \tag{13}$$

The discrete piecewise linear system defined in (11) is *completely observable* if and only if the total observability matrix (13) has rank equal to η (dimension of the state space) [13]. For sake of simplicity let us consider a scenario with two landmarks, in which the size of the error state vector is $\eta = 6(n+1) = 18$ (6 variables for each landmark plus 6 for the current robot pose). This simplification is made without loss of generality and the general case can be easily derived from the proposed analysis. Therefore, in this case, the TOM assumes the form:

$$\mathcal{O} = \begin{bmatrix} \overbrace{\quad}^{1} & \overbrace{\quad}^{2} & \overbrace{\quad}^{3} & \overbrace{\quad}^{4} & \overbrace{\quad}^{5} & \overbrace{\quad}^{6} \\ O_{6,6} & & \begin{bmatrix} -\hat{R}_1^\top & \hat{S}_1 \\ O_{3,3} & -\hat{R}_{\Delta_1}^\top \end{bmatrix} & & \begin{bmatrix} \hat{R}_1^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} & \\ O_{6,6} & & \begin{bmatrix} -\hat{R}_1^\top & M_{1,0}^1 \\ O_{3,3} & N_{1,0}^1 \end{bmatrix} & & \begin{bmatrix} \hat{R}_1^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} & \\ \vdots & & \vdots & & \vdots & \\ O_{6,6} & & \begin{bmatrix} -\hat{R}_1^\top & M_{1,0}^{\eta-1} \\ O_{3,3} & N_{1,0}^{\eta-1} \end{bmatrix} & & \begin{bmatrix} \hat{R}_1^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} & \\ \begin{bmatrix} -\hat{R}_2^\top & \hat{S}_2 \\ O_{3,3} & -\hat{R}_{\Delta_2}^\top \end{bmatrix} & & \begin{bmatrix} \hat{R}_2^\top & T_{2,0} \\ O_{3,3} & Q_{2,0} \end{bmatrix} & & O_{6,6} & \\ \begin{bmatrix} -\hat{R}_2^\top & M_{2,1}^1 \\ O_{3,3} & N_{2,1}^1 \end{bmatrix} & & \begin{bmatrix} \hat{R}_2^\top & T_{2,0} \\ O_{3,3} & Q_{2,0} \end{bmatrix} & & O_{6,6} & \\ \vdots & & \vdots & & \vdots & \\ \begin{bmatrix} -\hat{R}_2^\top & M_{2,1}^{\eta-1} \\ O_{3,3} & N_{2,1}^{\eta-1} \end{bmatrix} & & \begin{bmatrix} \hat{R}_2^\top & T_{2,0} \\ O_{3,3} & Q_{2,0} \end{bmatrix} & & O_{6,6} & \end{bmatrix}$$

with $T_{2,0} = -\hat{R}_2^\top \hat{R}_0 S_{u,1} \sum_{j=0}^{\eta-2} (R_{u,1}^\top)^j$ and $Q_1 = (R_{u,1}^\top)^{\eta-1}$. We can now study the observability property of our model by analyzing the column rank of $\mathcal{O}$. We can identify six block-columns, each one containing three columns. It is then possible to see that the third block column can be obtained as linear combination of the first and the fifth block-columns, hence the matrix is rank deficient, and the system is not completely observable. This is quite an intuitive result since we do not expect to be able to estimate absolute poses using only relative measurements: without imposing a reference frame only the relative poses can be observable. In pose-based EKF-SLAM it is common to assume the initial pose of the robot to be the reference frame for the mapping process. This assumption will be lately referred to as *grounding*, and the pose that is assumed to be known will be called *anchor*. The observability study can include the grounding process in two different ways: (i) including an absolute measurement of the initial pose, or (ii) marginalizing the sub-set of the state vector corresponding to the initial pose (i.e., the initial pose is assumed to be known and needs not be estimated by the EKF). We follow the first way, whereas the second approach can be verified to provide analogous results. An absolute measurement of pose at initial time corresponds to the following measurement matrix:

$$\mathcal{H}_0 = [O_{6,6} \;\; O_{6,6} \;\; I_{6,6}]$$

Considering this measurement, the TOM becomes:

$$\bar{\mathcal{O}} = \begin{bmatrix} O_{6,6} & O_{6,6} & I_{6,6} \\ \vdots & \vdots & \vdots \\ O_{6,6} & \begin{bmatrix} -\hat{R}_1^\top & \hat{S}_1 \\ O_{3,3} & -\hat{R}_{\Delta_1}^\top \end{bmatrix} & \begin{bmatrix} \hat{R}_1^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} \\ O_{6,6} & \begin{bmatrix} -\hat{R}_1^\top & M_{1,0}^1 \\ O_{3,3} & N_{1,0}^1 \end{bmatrix} & \begin{bmatrix} \hat{R}_1^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} \\ \vdots & \vdots & \vdots \\ O_{6,6} & \begin{bmatrix} -\hat{R}_1^\top & M_{1,0}^{\eta-1} \\ O_{3,3} & N_{1,0}^{\eta-1} \end{bmatrix} & \begin{bmatrix} \hat{R}_1^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} \\ \begin{bmatrix} -\hat{R}_2^\top & \hat{S}_2 \\ O_{3,3} & -\hat{R}_{\Delta_2}^\top \end{bmatrix} & \begin{bmatrix} \hat{R}_2^\top & T_{2,0} \\ O_{3,3} & Q_{2,0} \end{bmatrix} & O_{6,6} \\ \begin{bmatrix} -\hat{R}_2^\top & M_{2,1}^1 \\ O_{3,3} & N_{2,1}^1 \end{bmatrix} & \begin{bmatrix} \hat{R}_2^\top & T_{2,0} \\ O_{3,3} & Q_{2,0} \end{bmatrix} & O_{6,6} \\ \vdots & \vdots & \vdots \\ \begin{bmatrix} -\hat{R}_2^\top & M_{2,1}^{\eta-1} \\ O_{3,3} & N_{2,1}^{\eta-1} \end{bmatrix} & \begin{bmatrix} \hat{R}_2^\top & T_{2,0} \\ O_{3,3} & Q_{2,0} \end{bmatrix} & O_{6,6} \end{bmatrix}$$

In order to study the rank of the previous matrix let us consider the following matrix $\tilde{\mathcal{O}}$, obtained by selecting η rows from $\bar{\mathcal{O}}$:

$$\tilde{\mathcal{O}} = \begin{bmatrix} \begin{bmatrix} -\hat{R}_2^\top & \hat{S}_2 \\ O_{3,3} & -\hat{R}_{\Delta_2}^\top \end{bmatrix} & \begin{bmatrix} \hat{R}_2^\top & T_{2,0} \\ O_{3,3} & Q_{2,0} \end{bmatrix} & O_{6,6} \\ O_{6,6} & \begin{bmatrix} -\hat{R}_1^\top & \hat{S}_1 \\ O_{3,3} & -\hat{R}_{\Delta_1}^\top \end{bmatrix} & \begin{bmatrix} \hat{R}_1^\top & O_{3,3} \\ O_{3,3} & I_3 \end{bmatrix} \\ O_{6,6} & O_{6,6} & [I_{6,6}] \end{bmatrix}$$

Because of the block-triangular structure of the matrix $\tilde{\mathcal{O}}$, the determinant of $\tilde{\mathcal{O}}$ can be computed by multiplying the determinants of the blocks composing the main diagonal [16], i.e., $\det\tilde{\mathcal{O}} = \det(-\hat{R}_2^\top)\det(-\hat{R}_{\Delta_2}^\top)\det(-\hat{R}_1^\top)\det(-\hat{R}_{\Delta_1}^\top)\det(I_{6,6})$. Since the involved matrices are rotation or identity matrices we obtain $\det\tilde{\mathcal{O}} = 1$. The determinant is different from zero, then the matrix $\det\tilde{\mathcal{O}}$ is full rank; therefore, if a subset of η rows in $\bar{\mathcal{O}}$ are already linearly independent, also the rank of $\bar{\mathcal{O}}$ is η, and the system is completely observable. In conclusion, we demonstrated that, assuming a pose (the anchor) to be known, allows proving that the absolute poses included in the state vector in SLAM are completely observable from the relative pose measurements acquired in the time-interval $[1, k]$.

5. EXPERIMENTS AND DISCUSSION

In a real scenario, the computation of the relative pose measurements being input of the update phase of the EKF requires the following steps:
(1) *Information extraction*: salient elements are extracted from sensor data (e.g., visual features from cameras, positions of distinguishable objects, normal to walls, etc.);
(2) *Information synthesis*: the data extracted from 3D sensors are expressed in terms of 3D vectors (e.g., 3D points, normal to planes, line directions, etc.), each one described by a unique *descriptor*;
(3) *Data association*: the vectors extracted from the current observation are compared with the previous observations in order to establish the *correspondences*, i.e., the matches between each feature in the current observation and vectors occurring in previous measurements;
(4) *Vector Registration*: the relative pose (equivalently, the roto-translation) between pairs of poses from which two observations were made is retrieved, e.g., using a vector registration algorithm [8].

The concept of *descriptor* is purposely kept general. If the 3D vectors correspond to visual features, for instance, *computer vision* literature offers a large range of choices (e.g., SURT, SIFT [17]).

A crucial point in the implementation of the algorithm lies in *data association*. Since the descriptors are affected by uncertainty, it is often convenient to compare each vector in the current observation with all the possible matching candidates and select the optimal match according to some criterion (e.g., *double consistency check* for visual features). The set of candidate matches may be huge, since the features observed at current time need be compared with all the previous observations corresponding to landmarks in the state vector. When using vision sensors, a fast approach for detecting correspondences between current and past observations is the so

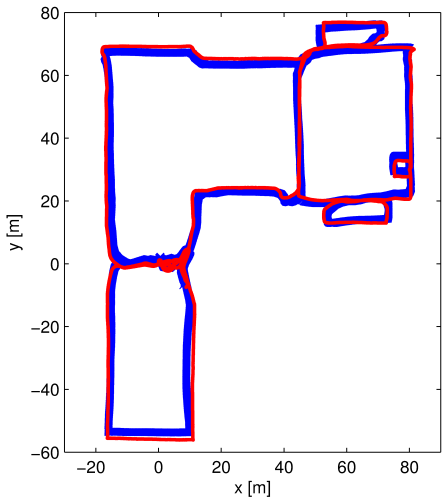

Figure 1. The ground truth trajectory from Rawseeds *dataset (*Bovisa 20090225b*) is shown as a red line, whereas the trajectory estimated with the proposed approach is plotted in blue.*

called *Vocabulary Tree*, proposed by Nister in [18]. A vocabulary tree is an appearance-based method for detecting similarity between two images. In our setup, we use the vocabulary tree as follows. As soon as a landmark is initialized in the pose-based architecture, the corresponding image is stored by the robot, and the image descriptor is computed. For detecting loop closing, the similarity between the current image and all the previous images is evaluated, by comparing the corresponding image descriptors. If the similarity score between the image acquired at time k and image corresponding to landmark j exceeds a given threshold, the j-th landmark is a good candidate for loop closing, i.e., it is likely that the robot is re-observing the landmark. Vocabulary tree approach is convenient in this context since it allows fast loop closing detection (the computation of the similarity scores requires few milliseconds also for large datasets [18]). Moreover, an online implementation is available [19].
Since the loop closing detection only accounts for appearance-based information, we found it convenient to improve robustness of the detection by including a geometric check within the EKF. This check was implemented using a standard *validation region approach* [10], in which measurements with high *normalized innovation squared* (NIS) are discarded.

We now have all the ingredients for presenting the real experiments. We tested the proposed framework on a real dataset from the *Rawseeds* project [20]. This allows using standard metrics for evaluating the approach in terms of estimation errors and correctness of the estimated 3D map. For the prediction phase we used the self-motion

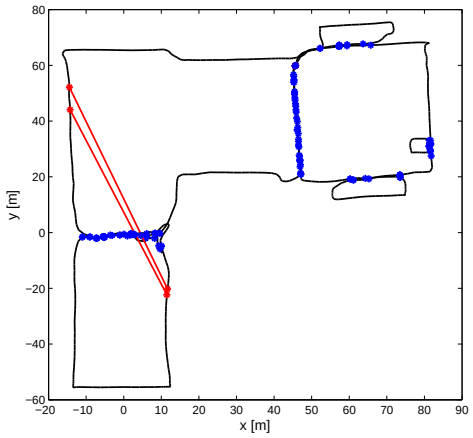

Figure 2. Estimated robot trajectory (black line). The blue and the red segments denote the candidate loop closings: the segments connect the poses at which the images matched by the vocabulary tree were acquired. The blue segments are the loop closings that passed the NIS test and were included in filter update, wheres the red segments were classified as outliers and discarded.

measurements from wheel odometry, whereas the update phase was based on the relative pose measurements from 3D vectors registration, where each vector was a visual feature (SURF) in a stereo pair. The trajectory estimated

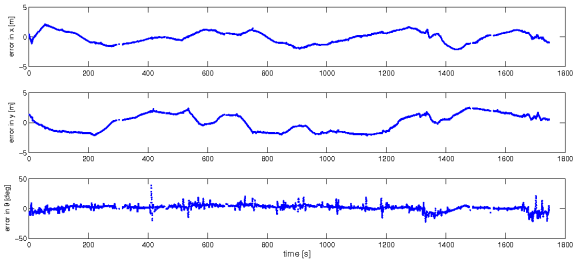

Figure 3. Cartesian error along the x-axis and the y-axis; orientation error in the yaw angle θ. The errors are reported as function of time, expressed in seconds.

with the proposed approach is shown as a blue line in Fig. 1. The corresponding ground truth trajectory (obtained using a *RTK*-GPS) is shown in red, in the same figure. Fig. 2 reports the candidate loop closings detected by the vocabulary tree and loop closings that passed the NIS test and that were actually included in the EKF-SLAM update phase.

A quantitative evaluation of the approach can be obtained from the statistics in Fig. 3. The figure reports the estimation errors, computed with the *metrics computation toolkit* available at [20]. In particular, they are shown the

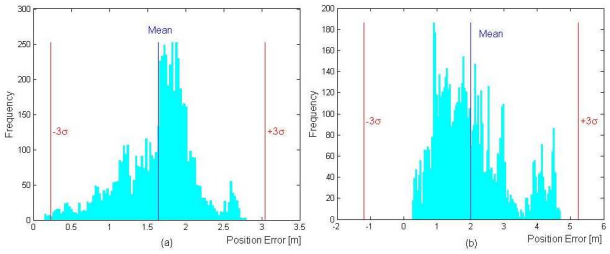

Figure 4. Histogram of Cartesian errors for the proposed approach (a) versus the CI-graph SLAM approach (b). The figure also reports the mean errors and the 3σ confidence intervals for the Cartesian errors.

Cartesian errors (considering the x and the y components of the error separately) and the orientation error (yaw angle θ) since the dataset corresponds to a planar problem. In Fig. 4(a), instead, it is shown a statistic on the Cartesian error over the whole experiment: the plot reports (i) the mean error, (ii) the histogram with the frequency with which a certain error occurs, (iii) the 3σ intervals that describe the region in which the error is likely to fall. We also report in Fig. 4(b) the corresponding error statistics of a benchmark solution of Rawseeds, that is based on standard EKF-SLAM and stereo data (stereo-based CI-graph SLAM [21]).

A further advantage of the considered framework is that pose estimation allows to retrieve a dense representation of the map. For instance we can obtain a dense point cloud representation from a stereo pair acquired at a given time step and we can use the corresponding pose estimate for roto-translating the dense point cloud, hence expressing all points in a common reference frame. The outcome of the mapping process is shown in Fig. 5 using the Rawseeds dataset.

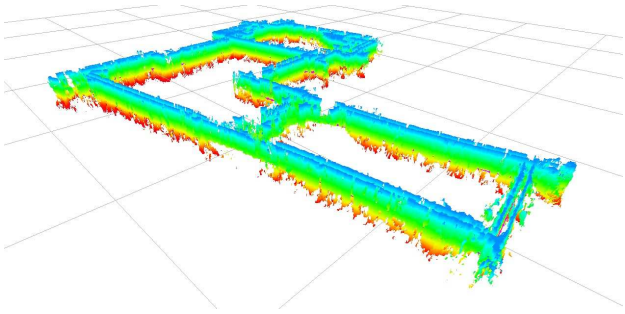

Figure 5. Dense 3D map of the scenario computed from the Rawseeds dataset.

6. CONCLUSION

In this work we presented a quaternion-based approach for solving Simultaneous Localization and Mapping from relative pose measurements. The approach does not use heuristics for forcing the unit norm constraint on quaternions of rotation and it is shown to be a high-level abstraction of several 3D perceptions of mobile robots. We also provided an observability analysis of the quaternion-based filter, showing that the SLAM system is completely observable over a given interval of time, as long as this interval includes an observation of the *anchor*, i.e., the landmark that was set to the origin of the reference frame for mapping. Finally, we reported experimental results for assessing accuracy and robustness of the proposed approach on real data.

REFERENCES

[1] J. Artieda, J.M. Sebastian, P. Campoy, J.F. Correa, I.F. Mondragon, C. Martinez, and M. Olivares, "Visual 3-D SLAM from UAVs," *Journal of Intelligent and Robotic Systems*, (55):4, 299-321, 2009.

[2] M. Bryson and S. Sukkarieh, "Observability analysis and active control for airborne SLAM," *IEEE Transactions on Aerospace and Electronic Systems*, (44):4, 261-280, 2008.

[3] H. Durrant-Whyte and T. Bailey, "Simultaneous Localization and Mapping (SLAM): Part I," *Robotics and Automation Magazine*, (13), 99-110, 2006.

[4] D.M. Cole and P.M. Newman, "Using laser range data for 3D SLAM in outdoor environments," *IEEE Int. Conf. on Robotics and Automation*, 1556-1563, 2006.

[5] R.M. Eustice, H. Singh, and J.J. Leonard, "Exactly sparse delayed-state filters for view-based SLAM," *Int. J. Robot. Res.*, (22):6, 1100-1114, 2006.

[6] V. Ila, J.M. Porta, J. Andrade-Cetto, "Information-based compact pose SLAM," *IEEE Trans. on Robotics*, (26):1, 78-93, 2010.

[7] P. Newman, "On the structures and solution of simultaneous localization and mapping problem," *Ph.D. thesis, Chalmers University of Technology and Gotëborg University*, 2007.

[8] L. Carlone and B. Bona, "On registration of uncertain three-dimensional vectors with application to robotics," *IFAC World Congress*, 2011.

[9] N. Trawny and S.I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," *University of Minnesota, Dept. of Comp. Sci. and Eng., Tech. Rep. 2005-002*, 2005.

[10] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan, "Estimation with applications to tracking and navigation," *John Wiley and Sons*, 2001.

[11] J.A. Castellanos, R. Martinez-Cantin, J.D. Tardós, and J. Neira, "Robocentric map joining: improving the consistency of EKF-SLAM," *Robotics and Autonomous Systems*, (55):1, 21-29, 2007.

[12] G. Huang, A.I. Mourikis, and S.I. Roumeliotis, "Observability-based rules for designing consistent EKF-SLAM estimators," *Int. J. Robot. Res.*, (29):5, 502-528, 2010.

[13] D. Goshen-Meskin and I.Y. Bar-Itzhack, "Observability analysis of piece-wise constant systems, Part 1: Theory," *IEEE Transactions on Aerospace and Electronic Systems*, (28):4, 1056-1067, 1992.

[14] J.H. Kim and S. Sukkarieh, "Improving the real-time efficiency of inertial SLAM and understanding its observability," *IEEE Int. Conf. on Intelligent RObots and Systems*, 1264-1269, 2004.

[15] F. Aghili, "3D simultaneous localization and mapping using IMU and its observability analysis," *Robotica*, (29):6, 805-814, 2010.

[16] R.A. Horn and C.R. Johnson, "Matrix analysis," *Cambridge University Press*, 1985.

[17] R.I. Hartley and A. Zisserman, "Multiple view geometry in computer vision," *Cambridge University Press*, 2000.

[18] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2161-2168, 2005.

[19] C. Cadena, D. Galvez-Lopez, F. Ramos, J.D. Tardós, and J. Neira, "Robust place recognition with stereo cameras," *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 5182-5189, 2010.

[20] S. Ceriani, G. Fontana, A. Giusti, D. Marzorati, M. Matteucci, D. Migliore, D. Rizzi, D.G. Sorrenti, and P. Taddei, "RAWSEEDS ground truth collection systems for indoor self-localization and mapping," *Autonomous Robots*, (27):34, 353-371, 2009.

[21] P. Piniés, L.M. Paz, and J.D. Tardós, "CI-Graph SLAM for 3D reconstruction of large and complex environments using a multicamera system," *International Journal of Field Robotics*, (27):5, 561-586, 2010.