2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

T-CoLoc: Leveraging Tethers for Reliable Co-Localization within an Underwater ROV Chain

J. Drupt[1], C. Dune[2], A. I. Comport[3], V. Hugel[2]

***Abstract*— Underwater Remotely Operated Vehicles (ROVs) exchange data with a control station *via* a communication cable. One or more intermediate robots can be placed along this tether to manage its shape and minimize the mechanical effects on the ROV. This work deals with the localization of a pair of underwater robots connected by a tether, in a previously unknown environment. While each robot can estimate its trajectory and a model of its surroundings using Simultaneous Localization And Mapping (SLAM) algorithms, aligning these observations in the same reference frame requires inter-robot data association. In this work, we introduce T-Coloc, a new method for aligning models' frames that leverages an estimation of the tether shape to align individual robot observations. An experimental validation in a pool demonstrates that T-CoLoc can align the trajectories of the two robots in the same reference frame with an error lower than 20 cm using the noisy shape estimation of a 3 m long tether.**

I. Introduction

The concept of a tethered, underwater robot chain was first introduced to reduce the mechanical disturbances exerted on underwater Remotely Operated Vehicles (ROVs) by their communication cables and to ensure system safety when exploring constrained and confined environments [1] (see Fig. 1).

While some works investigate tether shape modeling and estimation for underwater robot chains [2], [3], localizing a chain of underwater robots in an unknown environment is also a problem of underwater, multi-agent Simultaneous Localization and Mapping (SLAM). Underwater SLAM is typically based on sonar or cameras. Sonars are however not adapted to multiple vehicles operated close to each other due to acoustic interferences. Underwater visual conditions present unique challenges due to low textures, selective color absorption, backscattering, and the potential presence of embedded lights [4], [5]. Still, it has been shown that VSLAM can be a solution for close inspection of underwater structures provided the water is not overly turbid [6], [7], [8]. These reasons motivate the choice of VSLAM in this work.

Multi-agent VSLAM usually relies on place recognition to localize the multiple agents relative to each other. Still, multiple robots can explore the same area without even recording observations similar enough to perform place recognition. Additionally, underwater-specific visual conditions pose a particular challenge for visual feature matching and tracking, resulting in occasional non-detections of loop closures and VSLAM resets [8].

[1]LIRMM, University of Montpellier, CNRS, Montpellier, France
[2]COSMER Laboratory EA7398, Université de Toulon, France
[3]CNRS I3S Laboratory, Université Côte d'Azur, Sophia Antipolis, France

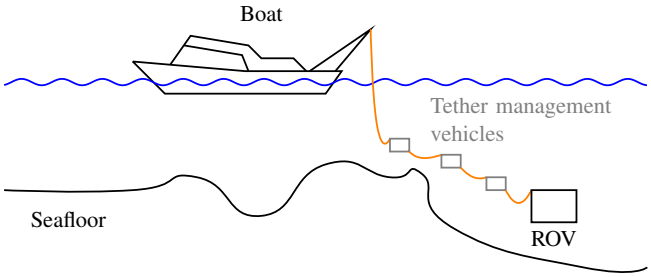


Fig. 1: Robot chain concept. The tether management vehicles compensate for mechanical effects of the cable on the leader ROV, and control its shape to keep it away from obstacles.

A unique feature of a robot chain is that its robots are physically connected, allowing their relative poses to be inferred from the shape of the tether segment between them. If the localization of one robot within a map is known — for example, using VSLAM — then the shape of the tether connecting that robot to its neighbor provides an estimate of the neighbor's position within the same reference frame. Iterating this process across pairs of robots links the entire chain to a common global frame.

This work investigates how the shape of the tether between two successive robots can be used as a cue for cooperative localization in an ROV pair. Assuming that each robot is equipped with a camera and performs VSLAM in its own local frame, the tether serves two purposes: (i) aligning the reference frames of the two local maps, and (ii) providing an initialization prior for robot relocalization in an existing map, for collaborative map construction or for recovering from a SLAM reset.

Section II presents the related works in underwater multi-robot SLAM and cable-based localization. The proposed cable-based co-localization method — namely T-CoLoc — is described in Section III. Experimental validations in a pool are depicted and analyzed in Section IV, before the conclusion and future works discussion in Section V.

II. Related work

Underwater SLAM is typically based on sonar and/or cameras, often combined with navigation sensors [9]. Underwater visual perception is challenged by multiple phenomena including water turbidity, suspended particles, backscattering and selective color absorption. Sonars are unaffected by these issues and can cover a large field of view (up to a few dozen meters). However, sonar images exhibit geometric ambiguities and have a high signal-to-noise ratio [10].

979-8-3315-8160-2/26/$31.00 ©2026 IEEE

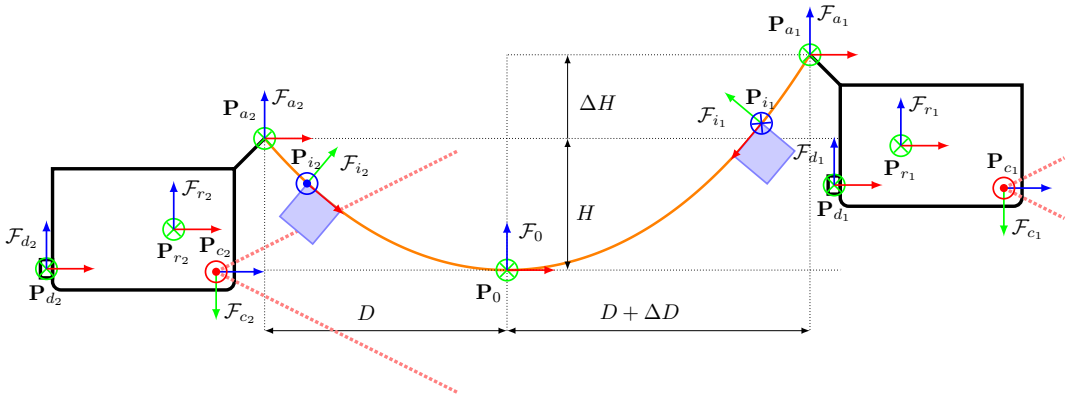
Cameras are usually preferred for close observation of underwater structures in clear water, and are the main feedback sensors used for remotely operated structure inspection and scientific exploration. In these conditions, cameras can provide rich, high-resolution information about their surroundings, allowing the use of off-the-shelf visual or visual-inertial SLAM algorithms [7]. ORB-SLAM3 [11] in particular has been identified as robust to underwater conditions by the literature [12], [8].

Most works on underwater multi-robot localization combine odometry measurements with occasional inter-robot acoustic ranging and bearing, using centralized or distributed probabilistic approaches. These works do not consider environment mapping [13], [14], [15]. Therefore, few works focus on underwater, multi-agent SLAM. [16] presents a distributed information filter-based SLAM framework for multiple Autonomous Underwater Vehicles (AUVs) equipped with a Doppler Velocity Log (DVL) and a Forward-Looking sonar (FLS). Neighboring AUVs exchange observed acoustic landmarks, which are merged under the assumption that all vehicles estimate their location with respect to the same reference frame. However, this hypothesis is quite strong in the absence of an absolute positioning reference underwater. In [17], a distributed graph-SLAM framework is used to fuse odometry and side-scan sonar measurements from multiple AUVs, and inter-vehicles acoustic ranging. The AUVs communicate local subgraphs including their pose and observed map features. Map observations from the multiple agents are merged using inter-robot positioning and sonar feature matching. Both of these works assume that the AUVs stay far enough apart to prevent sonar interference.

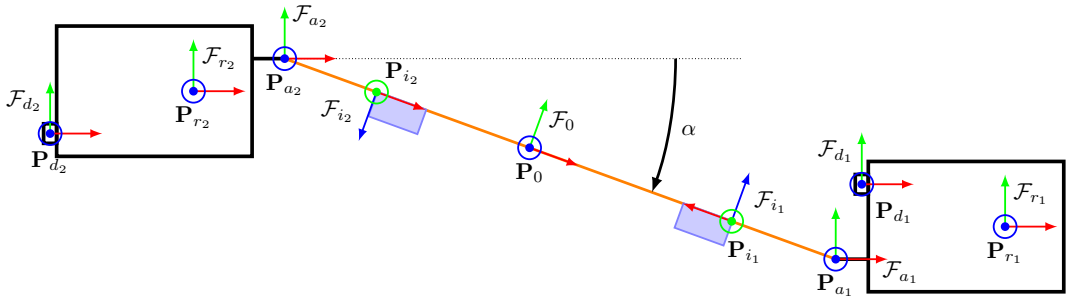
Unlike sonar-based methods, visual-based multi-agent SLAM works from the literature do not use inter-agent measurements to the best of the author's knowledge. Instead, observations from distinct agents are aligned using visual features matching [18], [19], [20]. [18] assumes that all agents can send their camera frames to a central server that estimates the map and the localizations of all vehicles, in a centralized manner. In practice, however, this is only feasible *via* a tether, as discussed in [20] which proposes a centralized, ORB-SLAM3-based SLAM for multiple ROVs operating in a small area. In [19], agents communicate only global image descriptors and features to the fusion center, instead of full images. Compression techniques for sharing images *via* acoustic communications are also investigated in [21]. All aforementioned works rely on map feature matching or place recognition to align incoming data from different agents. This assumes that the sensing modalities of the agents are similar enough to perform such an association and that the agents observe the same scene from a quite similar point of view. However, these assumptions may not be valid in practice. Another known limitation of visual SLAM in underwater environments is that the distortions in visual perception can sometimes cause failures in tracking map features [8]. The solution proposed in the literature is to restart the SLAM algorithm. The connection between the current and previous trajectory and map can be computed from place recognition, but remains unknown until the previously mapped area is revisited from a similar point of view [11].

An underwater robot chain is a multi-agent system in which robots are connected by tethers, which constraint their positioning. Some works in the literature examine cable-shape-based localization for tethered robotic systems. [2] considers the deployment of a mini-ROV from a larger AUV. The two vehicles are connected together by a specific 1-meter fiber optic cable, whose shape can be estimated using interferometry techniques to determine the ROV's position relative to the AUV. Without the need for such a specific cable, [3] proposes an affordable, easy-to-set-up ROV localization technique that uses local cable orientation measurements by IMUs, and exploits the piecewise linear shape of a cable equipped with a sliding ballast. Using the same principle, other underwater tether shape estimation methods can be adapted for ROV localization. One category of these works uses dynamic cable models [22], [23]. Still, these methods require precise knowledge of the system's physical parameters and have a prohibitive computational cost. Alternatively, a quasi-static model can be used. In the absence of currents, most works rely on the catenary model, a parametric curve that describes a non-extensible, homogeneous cable subject only to its weight and, optionally, buoyancy. In [1], a catenary shape is deduced from a visual feedback. However, their method is not very robust to visual disturbances. Instead, [24] proposes using a pair of IMUs placed near the endpoints of the tether. Cable sag is estimated with an accuracy of 2% of the cable's length, and the angle between the cable and the vehicle is estimated with an accuracy of 10°.

The current work introduces T-CoLoc, a new trajectory and map alignment method for a chain of underwater robots, considering it pair by pair. T-CoLoc leverages an estimation of the cable shape to align individual robot observations in the same reference frame. This is done by optimizing on a window of simultaneous SLAM-based pose estimations of each robot and the corresponding cable-based relative pose estimation. The key feature of this method is that it does not require to match exteroceptive observations from different robots. This means the robots do not need to observe the same scene from a quite similar point of view at any time. Another advantage of the proposed method is that if one of the robots experiences a SLAM failure that resets its map and trajectory estimate, the cable-based relative poses with respect to the other robot can be used to realign the new map and trajectory. Thus, continuity is ensured between the old and new localization and mapping of the failed robot without the need for place recognition.



(a) Side view



(b) Top view

Fig. 2: Frame definitions and 3D parameters.

III. T-CoLoc trajectories alignment

A. System, frames and notations

The system under study is represented in Fig. 2, and is composed of two robots tethered by a negatively buoyant cable of fixed length L. Each robot embeds a pressure sensor, an IMU and a camera, which is a minimal, usual sensor payload for small-size underwater robots.

For each robot $j \in \{1, 2\}$ we define the robot frame $\mathcal{F}_{r_j}$ with z-axis to the top of the robot and x-axis facing forward. $\mathcal{F}_{d_j}$ and $\mathcal{F}_{a_j}$ have the same orientation as $\mathcal{F}_{r_j}$ but with respective origins the robot's pressure sensors $\mathbf{P}_{d_j}$ and the attachment point of the tether $\mathbf{P}_{a_j}$. Each robot j embeds a monocular camera in $\mathbf{P}_{c_j}$, of camera frame $\mathcal{F}_{c_j}$. A pair of IMUs are placed on the cable in points $\mathbf{P}_{i_j}, j \in \{1, 2\}$, such that the x-axis of the inertial frames $\mathcal{F}_{i_j}$ are tangent to the cable, oriented downwards. In the following, we use the notation

$$ {}^{B}\mathbf{T}_{A} = \begin{bmatrix} {}^{B}\mathbf{R}_{A} & {}^{B}\mathbf{t}_{A} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \in SE(3) \tag{1} $$

to define the homogeneous rigid transformation from a frame $\mathcal{F}_A$ to a frame $\mathcal{F}_B$, with ${}^{B}\mathbf{R}_{A}$ a 3×3 rotation matrix and ${}^{B}\mathbf{t}_{A}$ a 3×1 translation vector.

B. Cable-based relative positioning

The tether is modeled by a catenary curve. By definition, this curve is contained in the vertical plane which contains attachment points $\mathbf{P}_{a_1}$ and $\mathbf{P}_{a_2}$. This shape is fully described by three parameters: $\{H, \Delta H, \alpha\}$ represented in Fig. 2. $H \in \mathbb{R}^+$ is the difference of vertical elevation between the tether's lowest point $\mathbf{P}_0$ and $\mathbf{P}_{a_2}$. $\Delta H \in \mathbb{R}$ is the signed difference of vertical elevation between $\mathbf{P}_{a_2}$ and $\mathbf{P}_{a_1}$. $\alpha \in]-\pi, \pi]$ is the yaw angle between Robot 2's forward axis (x-axis of $\mathcal{F}_{r_2}$) and the axis $\mathbf{P}_{a_2}\mathbf{P}_{a_1}$. These parameters are estimated using the method of [24], which uses the two IMUs placed on the cable and the robot's pressure sensors and embedded IMUs.

The homogeneous rigid transformation ${}^{a_2}\mathbf{T}_{a_1}$ can be deduced from $\{H, \Delta H, \alpha, {}^{w}\mathbf{R}_{r_1}, {}^{w}\mathbf{R}_{r_2}\}$ where ${}^{w}\mathbf{R}_{r_j} \in SO(3)$ is the orientation of robot $j \in \{1, 2\}$ measured by its embedded IMU with respect to the ENU frame. Let (ψ, θ, ϕ) be the Euler angles representation of ${}^{w}\mathbf{R}_{r_2}$ in the zyx intrinsic convention.

$$ {}^{w}\mathbf{R}_{r_2} = \mathbf{R}_z(\psi)\mathbf{R}_y(\theta)\mathbf{R}_x(\phi) \tag{2} $$

The position of point $\mathbf{P}_{a_1}$ in $\mathcal{F}_{a_2}$ is then:

$$ {}^{a_2}\mathbf{P}_{a_1} = \mathbf{R}_x(\phi)^T\mathbf{R}_y(\theta)^T \begin{bmatrix} d\cos(\alpha) \\ d\sin(\alpha) \\ \Delta H \end{bmatrix} \tag{3} $$

where:

$$ d = \frac{1}{C}(\operatorname{arccosh}(HC+1) + \operatorname{arccosh}((H+\Delta H)C+1)) \tag{4} $$

$$ C = \frac{2(2H + \Delta H + 2L\sqrt{H\frac{H+\Delta H}{L^2-\Delta H^2}})}{L^2 - (2H+\Delta H)^2} \tag{5} $$

Then,

$$ {}^{a_2}\mathbf{T}_{a_1} = \begin{bmatrix} {}^{w}\mathbf{R}_{r_2}^{T}\, {}^{w}\mathbf{R}_{r_1} & {}^{a_2}\mathbf{P}_{a_1} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \tag{6} $$

If each robot $j \in \{1, 2\}$ embeds a camera in $\mathbf{P}_{c_j}$, of camera frame $\mathcal{F}_{c_j}$ (see Fig. 2) the rigid transformation between $\mathcal{F}_{c_1}$ and $\mathcal{F}_{c_2}$ can be finally expressed as:

$$ {}^{c_2}\mathbf{T}_{c_1} = {}^{a_2}\mathbf{T}_{c_2}^{-1}\, {}^{a_2}\mathbf{T}_{a_1}\, {}^{a_1}\mathbf{T}_{c_1} \tag{7} $$

where ${}^{a_j}\mathbf{T}_{c_j}$ is fixed and known $\forall j \in \{1, 2\}$.

C. Individual VSLAM

Each robot runs a VSLAM algorithm. The pressure sensors are used to correct VSLAM scale and transfer the SLAM outputs to a fixed frame $\mathcal{F}_{w_j}, j \in \{1, 2\}$ with z-axis up and origin at the surface [25]. This way, one gets:

$$ {}^{w_1}\mathbf{T}_{w_2} = \left[\begin{array}{c|c} \mathbf{R}_z(\psi) & \begin{matrix} x \\ y \\ 0 \end{matrix} \\ \hline \mathbf{0}_{1\times 3} & 1 \end{array}\right] \tag{8} $$

where $\mathbf{R}_z(\psi)$ denotes a rotation of angle ψ around the z-axis. The z-coordinate of the translation component of the rigid transformation ${}^{w_1}\mathbf{T}_{w_2}$ is zero, since $\mathcal{F}_{w_1}$ and $\mathcal{F}_{w_2}$ both have their origin at the surface and z-axis up.

D. Trajectories alignment and relative pose refinement

Then, the SLAM outputs of the two robots can be aligned in the same reference frame by estimating parameters $\{\psi, x, y\}$ of equation (8). Let ${}^{w_j}\hat{\mathbf{T}}_{c_{j,k}}$ denote the estimate of the pose of camera $j \in \{1, 2\}$ according the the SLAM of robot j at time t_k, and ${}^{c_{2,k}}\hat{\mathbf{T}}_{c_{1,k}}$ denote the estimate of the pose of camera 1 with respect to camera 2 at time t_k. Using n samples of these transformations, we can write:

$$ {}^{w_1}\mathbf{T}_{w_2} = \underset{\{\psi,x,y\}\in\mathbb{R}^3}{\operatorname{argmin}} \sum_{k=0}^{n} || \left[\begin{array}{c|c} \mathbf{R}_z(\psi) & \begin{matrix} x \\ y \\ 0 \end{matrix} \\ \hline \mathbf{0}_{1\times 3} & 1 \end{array}\right] \ominus {}^{w_1}\hat{\mathbf{T}}_{w_2} ||^2 \tag{9} $$

where

$$ {}^{w_1}\hat{\mathbf{T}}_{w_2} = {}^{w_1}\hat{\mathbf{T}}_{c_{1,k}}\, {}^{c_{1,k}}\hat{\mathbf{T}}_{c_{2,k}}\, {}^{w_2}\hat{\mathbf{T}}_{c_{2,k}}^{-1} \tag{10} $$

As in [26], $\ominus$ defines the difference between two poses $\mathbf{T}_1, \mathbf{T}_2 \in SE(3)$, and is calculated in the tangent space of $SE(3)$ using the formula:

$$ \mathbf{T}_1 \ominus \mathbf{T}_2 := \mathfrak{log}(\mathbf{T}_1\, \mathbf{T}_2^{-1})^{\vee} \tag{11} $$

where $\mathfrak{log}(\cdot)$ is the logarithm map of $SE(3)$, which unwraps an element of the $SE(3)$ manifold to its tangent space, and $\cdot^{\vee}$ is the vee map of $\mathfrak{se}(3)$, which takes elements from the Lie algebra $\mathfrak{se}(3)$ to $\mathbb{R}^6$ [27] such that:

$$ {}^{\vee} : \mathfrak{se}(3) \longrightarrow \mathbb{R}^6 \tag{12} $$

$$ \begin{bmatrix} 0 & -\phi_z & \phi_y & \rho_x \\ \phi_z & 0 & -\phi_x & \rho_y \\ -\phi_y & \phi_x & 0 & \rho_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \longmapsto \begin{bmatrix} \rho_x \\ \rho_y \\ \rho_z \\ \phi_x \\ \phi_y \\ \phi_z \end{bmatrix} \tag{13} $$

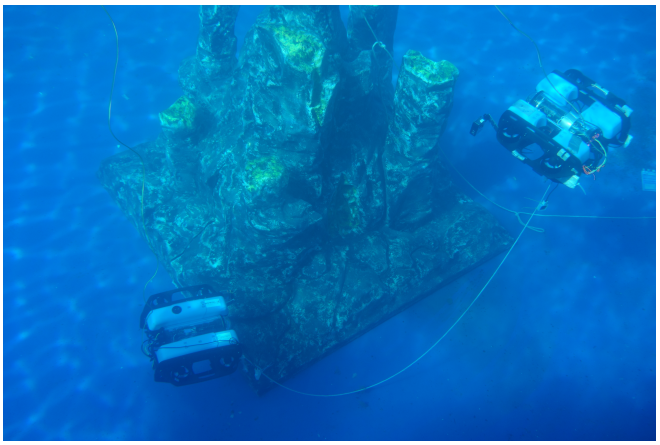


Fig. 3: Experimental pool setup: two BlueROV2 connected by a tether instrumented with IMUs, deployed in a 15 × 10 × 5 m pool containing a mock hydrothermal chimney.

Finally,

$$ {}^{w_1}\mathbf{T}_{w_2} = \underset{\{\psi,x,y\}\in\mathbb{R}^3}{\operatorname{argmin}} \sum_{k=0}^{n} ||\mathfrak{log}(\left[\begin{array}{c|c} \mathbf{R}_z(\psi) & \begin{matrix} x \\ y \\ 0 \end{matrix} \\ \hline \mathbf{0}_{1\times 3} & 1 \end{array}\right] {}^{w_1}\hat{\mathbf{T}}_{w_2}^{-1})^{\vee}||^2 \tag{14} $$

Equation (14) is solved using a Gauss-Newton on-manifold optimization of the x, y and ψ components [28]. In practice, the number of samples n should be small enough to ensure a fast optimization process and large enough to include as much information as possible in order to make a better estimate of ${}^{w_1}\mathbf{T}_{w_2}$.

Because the SLAM-estimated poses are significantly more accurate than the cable based ones, we also expect the transformations ${}^{w_1}\hat{\mathbf{T}}_{c_1}^{-1}\, {}^{w_1}\hat{\mathbf{T}}_{w_2}\, {}^{w_2}\hat{\mathbf{T}}_{c_2}$ to be a more accurate estimate of ${}^{c_{1,k}}\mathbf{T}_{c_{2,k}}$ than the cable-based estimate ${}^{c_{1,k}}\hat{\mathbf{T}}_{c_{2,k}}$.

IV. Experimental validation

The proposed cable-based SLAM frames alignment method — T-CoLoc — is evaluated in a pool, using a pair of ROVs connected by a mock tether.

A. Experimental set-up and methodology

1) Robotic set-up: The pool experiment is conducted in a 15 m large, 15 m long, 5 m depth sea water pool, using a pair of BlueROV2 tethered by a 3 m long, negatively buoyant cable. Each robot embeds a pressure sensor, an IMU and a fisheye camera. The cable is equipped with a pair of PhidgetsSpatial Precision 3/3/3 High Resolution IMUs placed 20 cm away from the closest ROV. This experimental set-up is shown in Fig. 3.

2) Reference trajectories: The robots are teleoperated to keep observing a mock hydrothermal chimney. These overlapping observations are used to generate an offline Structure from Motion (SfM) baseline *via* the Colmap software [29]. This baseline gives the trajectories of the two

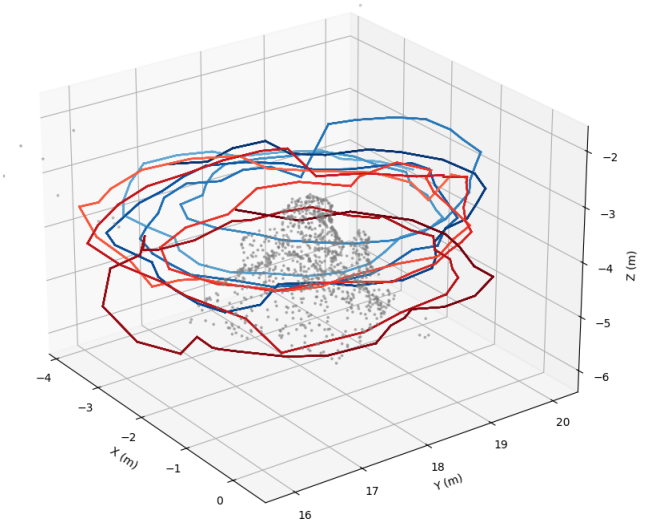


Fig. 4: Reference trajectories for Robots 1 (blue) and 2 (red), output by Colmap. The color scales indicate the direction of the trajectories: from lighter to darker colors. The grey points indicate the location of the mock hydrothermal chimney.

robots in the same world frame of known scale, and is used as a reference for evaluating localization accuracy [5]. These reference trajectories are represented in Fig. 4.

3) Test scenarios: In line with [12], [8], we choose monocular ORB-SLAM3 [11] as our VSLAM system. We noticed no SLAM failures in the test sequence. The tests are divided into two parts. First, we use the ORB-SLAM3 output for the entire sequence as is. Second, we simulate a SLAM failure by resetting Robot 2's SLAM at arbitrary points within the sequence.

B. Validation metrics

T-CoLoc is evaluated using the rigid transformations ${}^{w_1}\mathbf{T}_{c_2}$ and ${}^{c_2}\mathbf{T}_{c_1}$ estimated.

Accurately estimating ${}^{w_1}\mathbf{T}_{c_2}$ ensures that the robots' localizations are known in the same world frame, what is necessary for the robots to share map observations to avoid potential obstacles, plan paths, and interact with their environment in a real-world applications.

Conversely, accurate estimates of ${}^{c_2}\mathbf{T}_{c_1}$ are essential for inter-robot collision avoidance and to monitor the shape of the tether more accurately.

For ${}^{w_1}\mathbf{T}_{c_2}$ and ${}^{c_2}\mathbf{T}_{c_1}$, we compare:

- the poses given by T-CoLoc. Using the same notations as in Section III, these poses are respectively ${}^{w_1}\hat{\mathbf{T}}_{w_2}\,{}^{w_2}\hat{\mathbf{T}}_{c_2}$ and ${}^{w_1}\hat{\mathbf{T}}_{c_1}^{-1}\,{}^{w_1}\hat{\mathbf{T}}_{w_2}\,{}^{w_2}\hat{\mathbf{T}}_{c_2}$
- the poses given directly by combining Robot 1's VSLAM and the current tether-based measure of ${}^{c_1}\hat{\mathbf{T}}_{c_2}$. These poses are respectively ${}^{w_1}\hat{\mathbf{T}}_{c_1}\,{}^{c_1}\hat{\mathbf{T}}_{c_2}$ and ${}^{c_1}\hat{\mathbf{T}}_{c_2}$, using the notations from Section III. Since these estimates use the IMU-based tether-shape estimation of [24], they are labeled as 'T-IMU' in the following.
- the reference SfM transformations ${}^{w_1}\mathbf{T}_{c_2}$ and ${}^{c_2}\mathbf{T}_{c_1}$, noted 'Ref.' in the following.

We analyze the errors on each Degree Of Freedom (DOF) in $\{x, y, z, \phi, \theta, \psi\}$, where ϕ, θ, ψ are the Euler angles, defined with the same convention as in eq. (2). We also evaluate the error on the distance $d = \sqrt{x^2 + y^2 + z^2}$. For the translation DOFs and d, we use the error:

$$e_\lambda = |\lambda - \hat{\lambda}| \tag{15}$$

$\forall \lambda \in \{x, y, z, d\}$, where λ is the reference value and $\hat{\lambda}$ is the estimated value. For angular errors, we use the formula:

$$e_\theta = \text{atan2}(\sin(\theta - \hat{\theta}), \cos(\theta - \hat{\theta})) \tag{16}$$

$\forall \theta \in \{\phi, \theta, \psi\}$, where θ is the reference value and $\hat{\theta}$ is the estimated value.

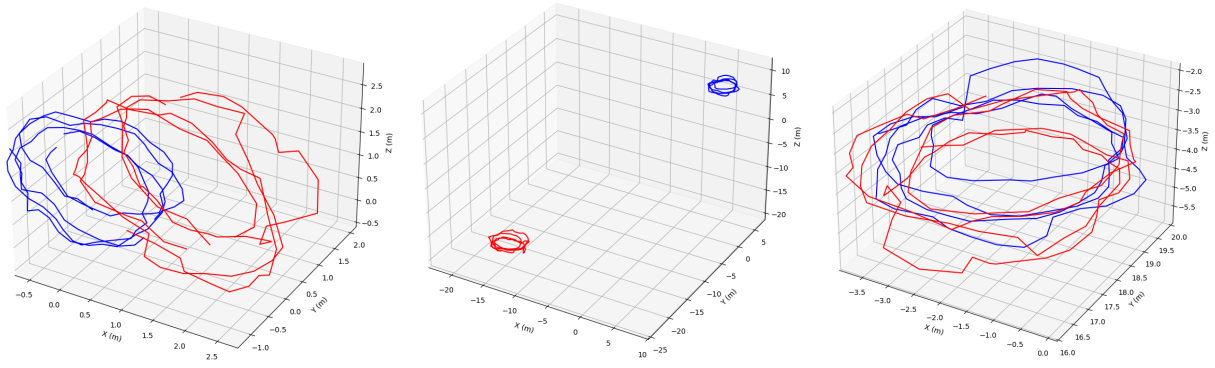
C. Cable-based alignment without SLAM reset

As ORB-SLAM3 did not encounter any failure while processing the sequence, we first evaluate T-CoLoc in the absence of SLAM resets using $n = 120$ pose samples, sampled every 3 s. Fig. 5 shows the trajectory alignment steps. The T-CoLoc and T-IMU estimates of ${}^{w_1}\mathbf{T}_{c_2}$ and ${}^{c_2}\mathbf{T}_{c_1}$ are compared to the SfM reference in Figures 6 and 7 respectively. Error statistics are presented in Table I.

TABLE I: DOF errors analysis for ${}^{w_1}\mathbf{T}_{c_2}$ and ${}^{c_2}\mathbf{T}_{c_1}$

		${}^{w_1}\mathbf{T}_{c_2}$		${}^{c_2}\mathbf{T}_{c_1}$	
DOF		T-CoLoc	T-IMU	T-CoLoc	T-IMU
e_x(m)	mean	**0.18**	0.56	**0.12**	0.47
	median	**0.17**	0.56	**0.13**	0.50
	σ	**0.05**	0.36	**0.07**	0.22
e_y(m)	mean	**0.10**	0.52	0.08	**0.07**
	median	**0.10**	0.48	0.07	**0.06**
	σ	**0.03**	0.34	**0.03**	0.06
e_z(m)	mean	0.08	**0.04**	**0.13**	0.61
	median	0.08	**0.03**	**0.15**	0.54
	σ	**0.02**	0.04	**0.06**	0.43
e_{roll}(m)	mean	**0.005**	0.03	**0.006**	0.03
	median	**0.004**	0.03	**0.004**	0.02
	σ	**0.004**	0.02	**0.01**	0.04
e_{pitch}(m)	mean	**0.005**	0.03	**0.02**	0.03
	median	**0.003**	0.03	**0.01**	0.05
	σ	**0.01**	0.04	**0.04**	0.06
e_{yaw}(m)	mean	**0.02**	0.07	**0.01**	0.03
	median	**0.006**	0.05	**0.004**	0.03
	σ	**0.04**	0.06	**0.02**	0.03
dist (m)	mean	**0.22**	0.86	**0.22**	0.86
	median	**0.21**	0.75	**0.21**	0.79
	σ	**0.05**	0.32	**0.03**	0.31

It can be observed by the error values of the T-CoLoc estimates are globally and significantly lower thant the T-IMU ones, for both ${}^{w_1}\mathbf{T}_{c_2}$ and ${}^{c_2}\mathbf{T}_{c_1}$. The T-CoLoc errors are up to 3 times lower for the translation DOFs, and up to 10 times lower for the orientation DOFs. This means that the proposed cable-based trajectory alignment efficiently improves the estimation of Robot 2's localization, not only in Robot 1's reference frame, but also with respect to Robot 1's mobile camera frame. Notably, the error on the distance between the robots is below 25 cm (which is also less than 10% of the cable's length). In practice, this means that,



(a) (b) (c)

Fig. 5: Alignment steps. Trajectories are represented in blue for Robot 1 and in red for Robot 2. Fig. 5a shows the initial scaled, misaligned trajectories output by ORB-SLAM3. Fig. 5b shows the ${}^{w_j}\hat{\mathbf{T}}_{c_j}, j \in \{1, 2\}$ trajectories, with scale, roll, pitch and vertical z-coordinate corrected and aligned using pressure measurements. In Fig. 5c, yaw and x and y-coordinates have been aligned in Robot 1's world frame $\mathcal{F}_{w_1}$ using the transformation ${}^{w_1}\mathbf{T}_{w_2}$ estimated *via* the tether's shape.

after this cable-based reference frame alignment has been done, the relative localization accuracy should by sufficient to efficiently prevent inter-vehicle collisions while the vehicles keep moving.

D. Cable-based alignment with SLAM reset

This test is based on a subsample of the pool sequence, during which we simulate a SLAM failure for Robot 2 at $t = 180$ s, followed by a SLAM reset at $t = 220$ s. The 40 s gaps simulates SLAM initialization failures. Robot 2's SLAM based trajectory and map are therefore decomposed in two disconnected parts, called parts A and B, which have different scales and reference frames, and are separated by a measurement gap.

Using T-CoLoc, we can align trajectory parts A and B to Robot 1's trajectory, what moves them to the same reference frame, with a metric scale. As in Section IV-C, we use pose samples taken every 3 s, resulting in $n = 58$ samples for trajectory part A and $n = 48$ samples for trajectory part B. Fig. 8 shows the 6 DOFs of the estimated trajectory ${}^{w_1}\mathbf{T}_{c_2}$ after alignment of the 2 sub-trajectories and maps.

One can see that trajectory parts A and B closely match the reference trajectory. These results demonstrate that T-CoLoc can effectively realign two parts of a SLAM output that are disconnected due to a SLAM failure and reset. This feature is particularly useful for ensuring continuity in localization and mapping so that pre-reset information remains usable. Based on this observation, T-CoLoc could also be used in future studies as an initial guess for pre- and post-reset feature matching and visual-based fine-tuning of the alignment.

V. Conclusion and Future Work

This work introduced T-CoLoc, a new method to align in the same reference frame the SLAM outputs of two underwater vehicles connected by a negatively buoyant tether. T-CoLoc exploits a coarse estimate of the tether's shape which uses local tangent measurements and robot's depth and orientation, released in [24].

Contrary to most multi-agent SLAM works, T-CoLoc does not need any inter-robot place recognition. That makes it available even if the robots never observe the same object with a similar point of view, but also if inter-robot sensor data matching is challenged due to the robots having different sensor types, or due to challenging perceptual conditions.

Experiments are conducted in pool with a pair of robots, using the visual-based SLAM framework ORB-SLAM3 [11]. We demonstrate that T-CoLoc can localize Robot 2 in Robot 1's SLAM reference frame with an accuracy about 10 to 20 cm and 1° for the translation and rotation degrees of freedom respectively. T-CoLoc also gives the relative localization of the robots with an accuracy below 20 cm and 1°. 20 cm is only 6% of the cable length, which is also the maximum distance between the robots. Consequently, a 20 cm accuracy can be considered small enough for collision avoidance. We also show that our cable-based alignment method can help coping with a SLAM reset for one of the robots by aligning previous and new SLAM estimates to the trajectory of the second robot.

In future works, the proposed alignment may be used as an initial guess to guide place recognition map features matching, for inter-robot map observation merging but also for realigning before-reset and after-reset SLAM outputs.

References

[1] M. Laranjeira, C. Dune, and V. Hugel, "Catenary-based visual servoing for tether shape control between underwater vehicles," *Ocean Engineering*, vol. 200, p. 107018, Mar. 2020.

[2] S.-C. Yu, J. Yuh, and J. Kim, "Armless underwater manipulation using a small deployable agent vehicle connected by a smart cable," *Ocean Engineering*, vol. 70, pp. 149–159, 2013.

[3] J. Drupt, C. Viel, C. Dune, and V. Hugel, "ROV localization using ballasted umbilical equipped with IMUs," *IEEE Journal of Oceanic Engineering*, vol. 50, no. 2, pp. 1045–1064, 2025.

[4] Y. Wang, W. Song, G. Fortino, L.-Z. Qi, W. Zhang, and A. Liotta, "An experimental-based review of image enhancement and image restoration methods for underwater imaging," *IEEE Access*, vol. 7, pp. 140 233–140 251, 2019.

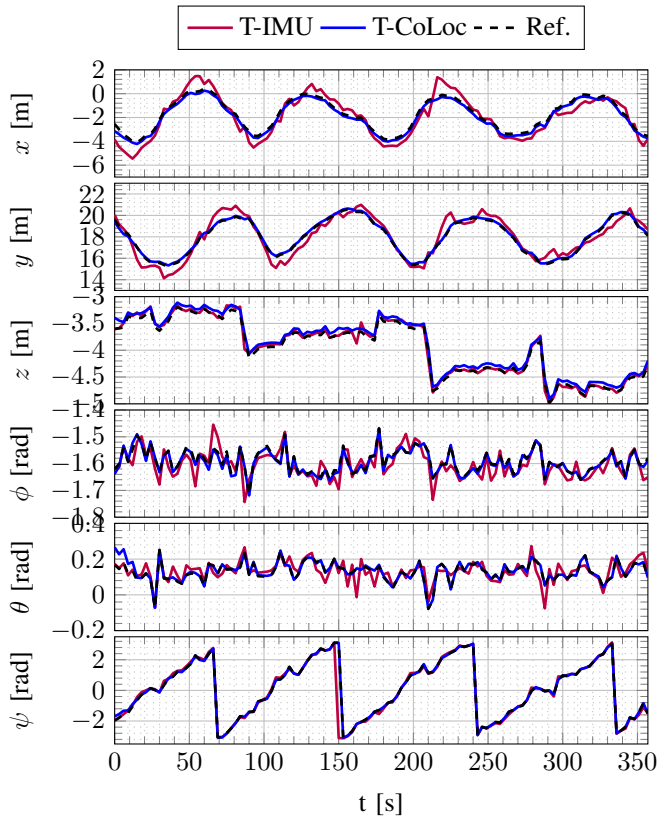


Fig. 6: Comparison of the six DOFs of T-IMU, T-CoLoc and reference values of ${}^{w_1}\mathbf{T}_{c_2}$.

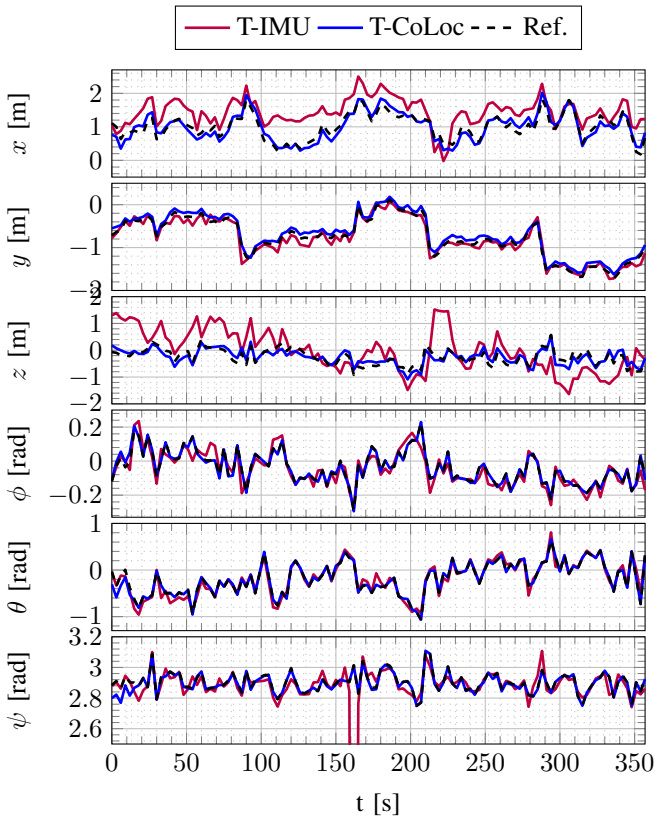


Fig. 7: Comparison of the six DOFs of T-IMU, T-CoLoc and reference values of ${}^{c_2}\mathbf{T}_{c_1}$.

[5] M. Ferrera, V. Creuze, J. Moras, and P. Trouvé-Peloux, "AQUALOC: An underwater dataset for visual-inertial-pressure localization," *The International Journal of Robotics Research*, vol. 38, no. 14, pp. 1549–1559, Oct. 2019.

[6] M. Ferrera, J. Moras, P. Trouvé-Peloux, V. Creuze, and D. Dégez, "The Aqualoc Dataset: Towards Real-Time Underwater Localization from a Visual-Inertial-Pressure Acquisition System," in *IROS Workshop - New Horizons for Underwater Intervention Missions: from Current Technologies to Future Applications*, Madrid, Spain, Oct. 2018.

[7] B. Joshi, S. Rahman, M. Kalaitzakis, B. Cain, J. Johnson, M. Xanthidis, N. Karapetyan, A. Hernandez, A. Q. Li, N. Vitzilaios, and I. Rekleitis, "Experimental comparison of open source visual-inertial-based state estimation algorithms in the underwater domain," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Nov. 2019, p. 7227–7233.

[8] J. Drupt, C. Dune, A. I. Comport, and V. Hugel, "Qualitative evaluation of state-of-the-art DSO and ORB-SLAM-based monocular visual slam algorithms for underwater applications," in *OCEANS 2023 - Limerick*, 2023, pp. 1–7.

[9] J. McConnell, I. Collado-Gonzalez, and B. Englot, "Perception for underwater robots," *Current Robotics Reports*, vol. 3, no. 4, p. 177–186, Oct. 2022.

[10] F. Yuan, F. Xiao, K. Zhang, Y. Huang, and E. Cheng, "Noise reduction for sonar images by statistical analysis and fields of experts," *Journal of Visual Communication and Image Representation*, vol. 74, p. 102995, Jan. 2021.

[11] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[12] Y. Li, X. Hong, J. Liu, Y. Chen, L. Xu, and C. Huang, "Implementation and evaluation of the ORB-SLAM3 in the underwater environment," in *International Conference on Electronic Information Engineering and Computer Communication (EIECC 2021)*, Z. Zhu and F. Cen, Eds. SPIE, May 2022, p. 77.

[13] L. Paull, M. Seto, and J. J. Leonard, "Decentralized cooperative trajectory estimation for autonomous underwater vehicles," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Sept. 2014, pp. 184–191.

[14] Z. J. Harris and L. L. Whitcomb, "Preliminary evaluation of cooperative navigation of underwater vehicles without a DVL utilizing a dynamic process model," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2018, p. 4897–4904.

[15] A. Quraishi, A. Bahr, F. Schill, and A. Martinoli, "A flexible navigation support system for a team of underwater robots," in *2019 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*. IEEE, Aug. 2019, p. 70–75.

[16] J. V. Diosdado and I. T. Ruiz, "Decentralised simultaneous localisation and mapping for AUVs," in *OCEANS 2007 - Europe*. IEEE, June 2007, p. 1–6.

[17] L. Paull, G. Huang, M. Seto, and J. J. Leonard, "Communication-constrained multi-AUV cooperative SLAM," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2015.

[18] S. Botelho, R. Neves, and L. Taddei, "Localization of a fleet of AUVs using visual maps," in *Europe Oceans 2005*. IEEE, 2005, pp. 1320–1325 Vol. 2.

[19] F. Bonin-Font and A. Burguera, "Towards multi-robot visual graph-slam for autonomous marine vehicles," *Journal of Marine Science and Engineering*, vol. 8, no. 6, p. 437, June 2020.

[20] J. Drupt, A. I. Comport, C. Dune, and V. Hugel, "MAM^{3}SLAM: Towards underwater-robust multi-agent visual slam," *Ocean Engineering*, vol. 302, p. 117643, June 2024.

[21] F. Bonin-Font and A. B. Burguera, "Image compression for underwater multi-robot loop closing," in *OCEANS 2023 - Limerick*. IEEE, June 2023, p. 1–7.

[22] S. M. Hong, K. N. Ha, and J.-Y. Kim, "Dynamics modeling and motion

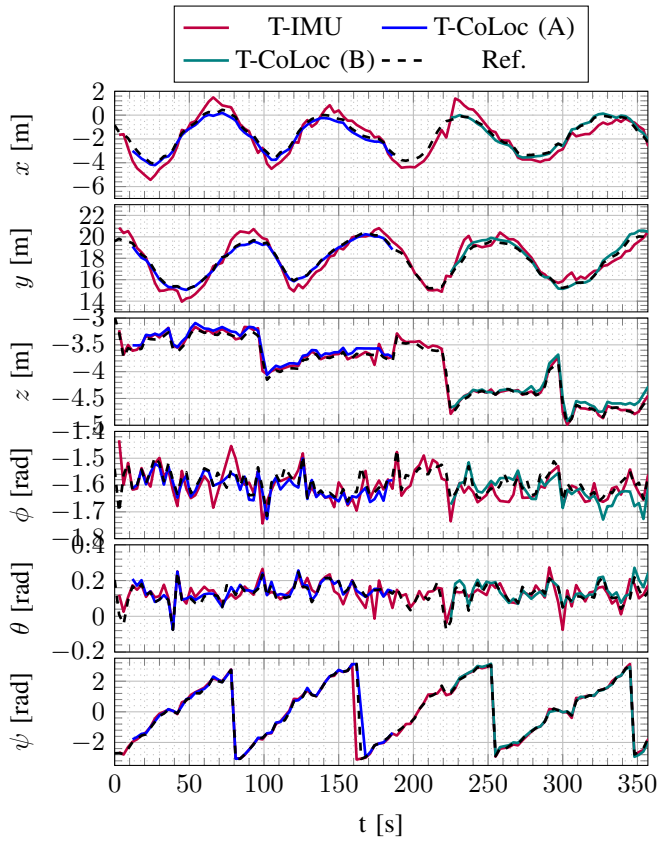


Fig. 8: Comparison of the six DOFs of T-IMU, T-CoLoc and reference values of ${}^{w_1}\mathbf{T}_{c_2}$. T-CoLoc trajectories for parts A and B are displayed in blue and teal respectively.

simulation of USV/UUV with linked underwater cable," *Journal of Marine Science and Engineering*, vol. 8, no. 5, p. 318, Apr. 2020.

[23] Y. Meng, X. Xu, and M. Zhao, "Dynamics calculation of complex deep-sea cable system based on hybrid optimization algorithm," *Ocean Engineering*, vol. 200, p. 107041, Mar. 2020.

[24] J. Drupt, C. Dune, A. I. Comport, S. Seillier, and V. Hugel, "Inertial-measurement-based catenary shape estimation of underwater cables for tethered robots," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 6867–6872.

[25] C. Boittiaux, "Visual localization for deep-sea long-term monitoring," Ph.D. dissertation, Université de Toulon, 2023.

[26] H. Strasdat, "Local accuracy and global consistency for efficient visual slam," Ph.D. dissertation, Imperial College of London, 2012.

[27] F. Dellaert and M. Kaess, *Factor Graphs for Robot Perception*. IEEE, 2017.

[28] T. D. Barfoot, *State estimation for Robotics*. Cambridge University Press, 2017.

[29] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.