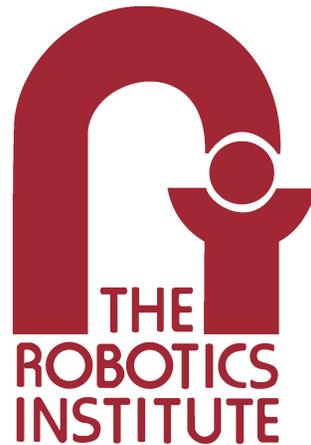


Underwater AprilTag SLAM and calibration for high precision robot localization

Eric Westman and Michael Kaess

October 2018



Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
CMU-RI-TR-18-43

Abstract

In this work we present a SLAM framework using the popular AprilTag fiducials for obtaining precise, drift-free pose estimates of an underwater vehicle. The framework also allows for simultaneous calibration of extrinsics between the camera and the vehicle odometry coordinate frame. The pose estimates may be used for various underwater tasks such as mapping and inspection, or as a ground-truth trajectory to evaluate the accuracy of other localization methods. We evaluate the effectiveness of the system with real-world experiments in a test-tank environment and demonstrate that it corrects drift that accumulates with dead-reckoning localization.

1 Introduction and background

Many underwater robotic tasks require high-precision vehicle localization. Vehicle odometry may be measured by an inertial measurement unit (IMU) or a Doppler velocity log (DVL), among other sensors. However, these pose estimates will drift unboundedly with time, as they rely on dead reckoning (integration of odometry measurements). For traditional non-underwater robotics, ground-truth trajectories of robots or sensors are typically acquired by a camera-based motion capture system or laser surveying equipment. While such motion capture systems exist for underwater tracking [15], the inherent difficulties presented by underwater optics and electronics make such systems cost-prohibitive for many applications. Furthermore, these systems are only practically usable in a controlled laboratory setting, and not in the field. We aim to provide a localization solution that:

- corrects drift that accumulates with dead-reckoning and bounds the pose uncertainty
- incorporates any localization information from multiple on-board sensors
- automatically solves for the extrinsics between the camera and odometry coordinate frames
- is significantly less costly than an underwater motion capture system
- is highly reconfigurable and requires minimal labor to setup and operate in the laboratory and in the field.

Various visual SLAM systems have proven capable of satisfying all of the above requirements, both in standard open-air environments as well as underwater, with various limitations and precision [7]. In uncontrolled environments, natural features are often detected and used as the landmarks in a SLAM formulation. A variety of feature descriptors have been formulated in order to perform the critical task of *data association*: matching features locally for feature tracking or globally for loop closure [2]. While descriptors aid the matching process, outlier rejection algorithms such as RANSAC usually must be employed to reject incorrect correspondences. Nevertheless, incorrect feature correspondences may persist and negatively affect the SLAM result.

Visual fiducials (easily identifiable, artificial markers placed in the robot’s environment) are often used to provide strong features and a robust solution to the data association problem. Various types of fiducials and detection methods have been proposed in recent years and have been widely used in the field of robotics [5, 13, 18]. In our proposed visual SLAM framework, we utilize the AprilTag system [13], as it provides particularly robust data association correspondences and can even identify partially occluded fiducials. Since the proposed algorithm performs online mapping, the fiducials may be placed anywhere in the environment, as long as they remain stationary. However, placing the AprilTags where they will be viewed most frequently over the course of the vehicle’s mission will help the SLAM algorithm to generate the best localization, mapping, and calibration results.

The proposed system can accommodate the use of individual AprilTag fiducials as the landmarks in the SLAM system as well as our custom-made AprilTag boards, which are shown in Figure 2. We printed four AprilTags of the same size in a square pattern on each aluminum dibond board. This aids the SLAM process by reducing the number of degrees of freedom that need to be estimated. If eight AprilTags are utilized in the form of two boards, then only 12 DOF must be estimated for the landmarks (two 6-DOF poses) in contrast to the 48 DOF that would be required to model the poses of eight individual AprilTags. In the remainder of this work, we will describe our system as it pertains to our custom-made AprilTag boards.

Several recent works have proposed using visual fiducials to achieve high-precision, drift-free localization of an underwater vehicle using an EKF framework [8] and a particle filter framework [9]. These systems have proven successful at reducing localization error over dead-reckoning, but do not allow for simultaneously solving for the camera extrinsics (relative to odometry coordinate frame), and presumably utilize manual extrinsics measurements. We do not explicitly compare our localization method to these in this work, as each system is highly tailored to the specific underwater vehicle and testing environment.

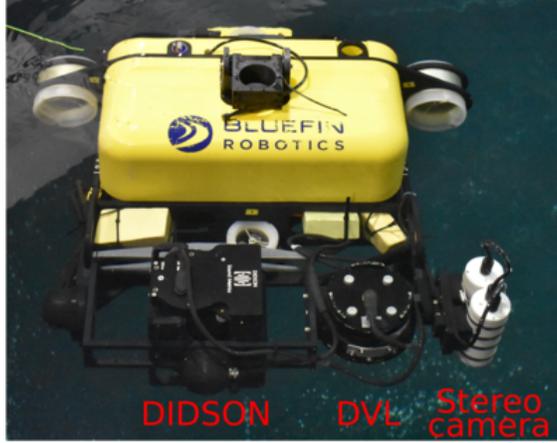


Figure 1: The Bluefin HAUV which is used in our real-world experiments. The DVL and stereo camera are mounted in front of the vehicle, and fixed facing downward at the bottom of the tank. The DIDSON imaging sonar is not used in this SLAM system.

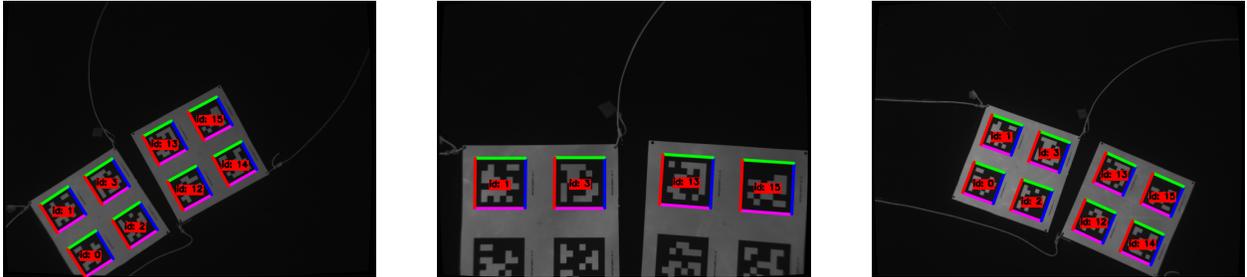


Figure 2: Sample images from one of our test tank datasets, showing the AprilTags detected from various vehicle depths and viewing angles.

2 Vehicle configuration

While the proposed framework may be used with a variety of underwater robots, we perform our experiments with the Bluefin Hovering Autonomous Underwater Vehicle (HAUV) [6], as shown in Figure 1. This vehicle is equipped with several sensors for onboard navigation: a 1.2MHz Teledyne/RDI Workhorse Navigator Doppler velocity log (DVL), an attitude and heading reference system (AHRS), and a Paroscientific Digiquartz depth sensor. The AHRS utilizes a Honeywell HG1700 IMU to measure acceleration and rotational velocities. The DVL is an acoustic sensor that measures translational velocity with respect to the water column or a surface, such as the seafloor, test tank floor, or ship hull.

A proprietary navigation algorithm fuses measurements from these sensors to provide odometry estimates for the vehicle, in the coordinate frame of the DVL sensor. It is important to note that the depth sensor gives direct measurements of the vehicle’s depth, or Z position. The AHRS is also capable of producing very accurate, drift-free estimates of the vehicle’s pitch and roll angles by observing the direction of gravity. The X and Y translation and yaw rotation are not directly observable, and are therefore usually estimated by dead reckoning of the DVL and IMU odometry measurements. The pose estimate will inevitably drift in these directions over long-term operation.

This naturally leads to a formulation that treats each of the vehicle’s odometry measurements as two separate types of constraints: a relative pose-to-pose constraint on XYH motion (X and Y translation, and heading, or yaw, rotation) and a unary ZPR constraint (Z position, pitch and roll rotation) [16]. This correctly models the dead reckoning pose estimates as directly observable and drift-free in the ZPR directions and unboundedly drifting in the XYH directions.

A stereo pair of two Prosilica GC1380 cameras are also mounted on the vehicle, on the same roll cage as the DVL sensor. Although our system accommodates utilizing images from both cameras in stereo configuration, in this work we will describe the use of a single camera for monocular SLAM. The camera intrinsics are calibrated underwater using the pinhole camera model, after correcting the images for radial and tangential distortion, as described in [17].

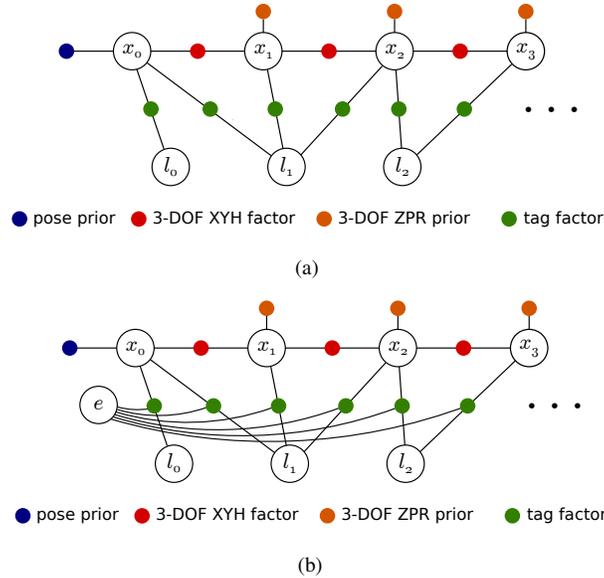


Figure 3: (a) Factor graph representation of AprilTag SLAM (b) Factor graph representation of AprilTag SLAM with automatic calibration of vehicle-camera extrinsics. A node is represented as a large, uncolored circle labeled with the corresponding variable name. A factor is represented as a small, colored circle, with the color corresponding to the category of measurement, as indicated by the legend. Factor labels are omitted for simplicity.

3 Proposed SLAM algorithm

3.1 Factor-graph SLAM

We model the SLAM problem with a factor-graph formulation, which is shown in Figure 3a. A factor graph is a bipartite graph with two types of vertices: nodes that represent the *variables* in the optimization and factors that represent the *measurements* that provide constraints. The edges from factors to nodes describe the dependency structure of the optimization: the nodes connected to a factor are constrained by that factor’s measurement.

In the standard SLAM formulation, the variables consist of the poses in the trajectory and the observed landmarks. The state is comprised of all of the variables: $\Theta = \{x_1, \dots, l_1, \dots\}$, where x_i is the i th vehicle pose and l_j is the j th landmark pose. Note that we explicitly model the vehicle poses rather than camera poses in the factor graph. However, the camera poses may be computed using the vehicle-camera extrinsics. Here a landmark represents one of our custom-made boards that has four AprilTag fiducials printed in precisely known locations. The measurement vector is comprised of all measurements $\mathcal{Z} = \{r_0, u_1, \dots, v_1, \dots, m_1, \dots\}$, where u_i is an XYH odometry measurement that constrains poses x_{i-1} and x_i , v_i is a ZPR measurement that constrains pose x_i , and m_k is an observation of any AprilTag. Finally, we denote with r_0 a prior measurement placed on the first pose to tie down the trajectory to a global coordinate frame.

3.2 Factor-graph SLAM with extrinsics calibration

The typical SLAM formulation presented in the previous sub-section assumes the camera-vehicle extrinsics are known *a priori* and are treated as constant in the optimization. However, these extrinsics may be very difficult to measure precisely by hand. However, we may incorporate this variable seamlessly into the factor-graph formulation, as shown graphically in Figure 3b. We simply add the extrinsics e as an additional 6-DOF pose to the state, so that $\Theta = \{x_1, \dots, l_1, \dots, e\}$. This variable is then constrained by the AprilTag measurement factors, but not by any odometry factors, since the extrinsics describe a relative pose from the vehicle coordinate frame to the camera coordinate frame. Note that we assume this transformation is constant throughout the entire operating sequence of the vehicle (the camera is fixed relative to the odometry frame).

3.3 Nonlinear least-squares optimization

The optimization typically used with the factor-graph representation is maximum a posteriori (MAP), which attempts to maximize the probability of the variables given the measurements. This optimization may be written as

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} p(\Theta | \mathcal{Z}) \quad (1)$$

$$= \underset{\Theta}{\operatorname{argmax}} p(\Theta) p(\mathcal{Z} | \Theta) \quad (2)$$

$$= \underset{\Theta}{\operatorname{argmax}} p(x_0) \prod_{i=1}^n p(u_i | x_{i-1}, x_i) p(v_i | x_i) \prod_{k=1}^m p(m_k | x_{i_k}, l_{j_k}, e). \quad (3)$$

In Equation 3 we use the factorization of the joint likelihood of the measurements encoded by the factor-graph. The prior factor r_0 materializes as $p(x_0)$. The tag measurement likelihood $p(m_k | x_{i_k}, l_{j_k}, e)$ and odometry measurement likelihood $p(u_i | x_{i-1}, x_i)$ are conditioned on particular values of the variables involved in the measurement. Here we use l_{j_k} and x_{i_k} to denote the board observed by m_k and the pose from which the observation was made, respectively. See [3] for additional details about the factor-graph based SLAM framework.

The MAP estimation is typically performed under the assumption that all measurements and priors are normally distributed random variables:

$$p(x_0) = \mathcal{N}(x_{r_0}, \Sigma_0) \quad (4)$$

$$p(u_i | x_{i-1}, x_i) = \mathcal{N}(f(x_{i-1}, x_i), \Gamma_i) \quad (5)$$

$$p(v_i | x_i) = \mathcal{N}(g(x_i), \Lambda_i) \quad (6)$$

$$p(m_k | x_{i_k}, l_{j_k}, e) = \mathcal{N}(h(x_{i_k}, l_{j_k}, e), \Xi_k). \quad (7)$$

Here we use $f(x_{i-1}, x_i)$, $g(x_i)$ and $h(x_{i_k}, l_{j_k}, e)$ to denote the *measurement functions* (also called *prediction functions*) for u_i , v_i , and m_k , respectively, and x_{r_0} is the specified prior pose estimate. Making the assumption of normally distributed measurements allows the MAP optimization to be simplified as a nonlinear least-squares optimization:

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} -\log \left[p(x_0) \prod_{i=1}^n p(u_i | x_{i-1}, x_i) p(v_i | x_i) \prod_{k=1}^m p(m_k | x_{i_k}, l_{j_k}, e) \right] \quad (8)$$

$$= \underset{\Theta}{\operatorname{argmin}} \|x_{r_0} \ominus x_0\|_{\Sigma_0}^2 + \sum_{i=1}^n \left(\|u_i - f(x_{i-1}, x_i)\|_{\Gamma_i}^2 + \|v_i - g(x_i)\|_{\Lambda_i}^2 \right) + \sum_{k=1}^m \|m_k - h(x_{i_k}, l_{j_k}, e)\|_{\Xi_k}^2 \quad (9)$$

where we use the common notation for Mahalanobis distance: $\|v\|_{\Sigma}^2 = v^T \Sigma^{-1} v$, and \ominus denotes the logmap difference between the two manifold elements upon which it operates. This nonlinear least-squares formulation allows for efficient incremental inference using state of the art algorithms that can produce real-time state estimates depending on the size and sparsity of the system [10, 11].

3.4 Measurement functions

3.4.1 6-DOF pose prior factor

The operator \ominus for two elements of the $SE(3)$ Lie group, such as in the 6-DOF prior factor on x_0 , is the logarithm map of the relative transformation between the elements:

$$x_{r_0} \ominus x_0 = \log(x_{0, r_0}). \quad (10)$$

Here we use $x_{a,b}$ to denote the relative transformation from pose x_a to x_b . The explicit form of the $SE(3)$ logarithm map and exponential map, which is used to update the 6-DOF poses in optimization, are detailed in [1].

3.4.2 XYH and ZPR factors

For the XYH and ZPR measurement factors, we make use of the Euler angle representation of 3D rotations to represent a pose x_i as the 6-vector $[\psi_{x_i}, \theta_{x_i}, \phi_{x_i}, t_{x_i}^x, t_{x_i}^y, t_{x_i}^z]^T$, where $\psi_{x_i}, \theta_{x_i}, \phi_{x_i}$ and are the roll, pitch, and yaw (heading) angles, respectively. Since the measurements u_i and v_i consist of only 3-DOF each, we represent them using the appropriate Euler angles and translation components:

$$u_i = [t_{u_i}^x \quad t_{u_i}^y \quad \phi_{u_i}]^T \quad (11)$$

$$v_i = [t_{v_i}^z \quad \theta_{v_i} \quad \psi_{v_i}]^T. \quad (12)$$

Likewise, the prediction functions $f(x_{i-1}, x_i)$ and $g(x_i)$ are comprised of the corresponding components of $x_{i-1,i}$, the relative transformation from x_{i-1} to x_i , respectively:

$$f(x_{i-1}, x_i) = [t_{x_{i-1,i}}^x \quad t_{x_{i-1,i}}^y \quad \phi_{x_{i-1,i}}]^T \quad (13)$$

$$g(x_i) = [t_{x_i}^z \quad \theta_{x_i} \quad \psi_{x_i}]^T. \quad (14)$$

All Euler angles are normalized to the range $[-\pi, \pi)$ radians.

3.4.3 Tag measurement factors

A tag measurement m_k is an 8-vector comprised of the (u, v) pixel coordinates of the four corners of a detected AprilTag:

$$m_k = [c_{k,1}^u \quad c_{k,1}^v \quad c_{k,2}^u \quad c_{k,2}^v \quad c_{k,3}^u \quad c_{k,3}^v \quad c_{k,4}^u \quad c_{k,4}^v]^T. \quad (15)$$

Since the side length s and placement of each AprilTag on its board is known *a priori*, the 3D corner points are treated as constants relative to the board's coordinate frame. The origin of the board's frame is at the center of the board and the x and y axes are parallel to the sides of the AprilTags. The homogeneous 3D coordinates of AprilTag corner i in the board frame corresponding to the measurement m_k is denoted as $d_{k,i} = [d_{k,i}^x \quad d_{k,i}^y \quad 0 \quad 1]^T$. Using the pinhole camera model with a calibrated intrinsics matrix K , we define q_k as the relative transformation from l_{j_k} to $x_{i_k} * e$. That is, q_k is the pose of the camera at timestep i_k relative to the frame of AprilTag l_{j_k} . Under this projection model, the tag measurement function is:

$$h(x_{i_k}, l_{j_k}) = \Omega(KT(q_k) [d_{k,1} \quad d_{k,2} \quad d_{k,3} \quad d_{k,4}]) \quad (16)$$

where the function $\Omega(\cdot)$ normalizes each homogeneous 3-vector column of the input matrix and reshapes the entire matrix to an 8-vector, in the form of m_k . $T(q_k)$ creates the 4×4 transformation matrix that corresponds to the 6-DOF relative pose q_k .

3.5 Noise models

As described in Section 3.3, we use Gaussian noise models for all variables. The covariance matrices are diagonal square matrices with size equal to the dimension of the corresponding measurement:

$$\Gamma_i = \text{diag}([\sigma_{i,x}^2 \quad \sigma_{i,y}^2 \quad \sigma_{i,\phi}^2]) \quad (17)$$

$$\Lambda_i = \text{diag}([\sigma_{i,z}^2 \quad \sigma_{i,\theta}^2 \quad \sigma_{i,\psi}^2]) \quad (18)$$

$$\Xi_k = \sigma_c^2 I_8 \quad (19)$$

where $\text{diag}(v)$ creates a diagonal matrix with vector v along the diagonal. Σ_0 is set to an arbitrarily small multiple of I_6 to tie down the trajectory to a global reference frame. The specific values of the noise model parameters are shown in Table 1.

Noise models

	$\sigma_{i,x}$	$\sigma_{i,y}$	$\sigma_{i,\phi}$	$\sigma_{i,z}$	$\sigma_{i,\theta}$	$\sigma_{i,\psi}$	σ_c
Value	$0.005 + 0.002dt$	$0.005 + 0.002dt$	$0.005 + 0.002dt$	0.02	0.005	0.005	1
Units	m	m	rad	m	rad	rad	pixels

Table 1: Values of noise model parameters for the XYH and ZPR odometry factors and AprilTag measurement factors. dt denotes the difference in time between the two poses involved in the corresponding XYH odometry factors. The uncertainty of the ZPR measurements is constant due to direct observability, and the uncertainty of the XYH factors increases linearly over time.

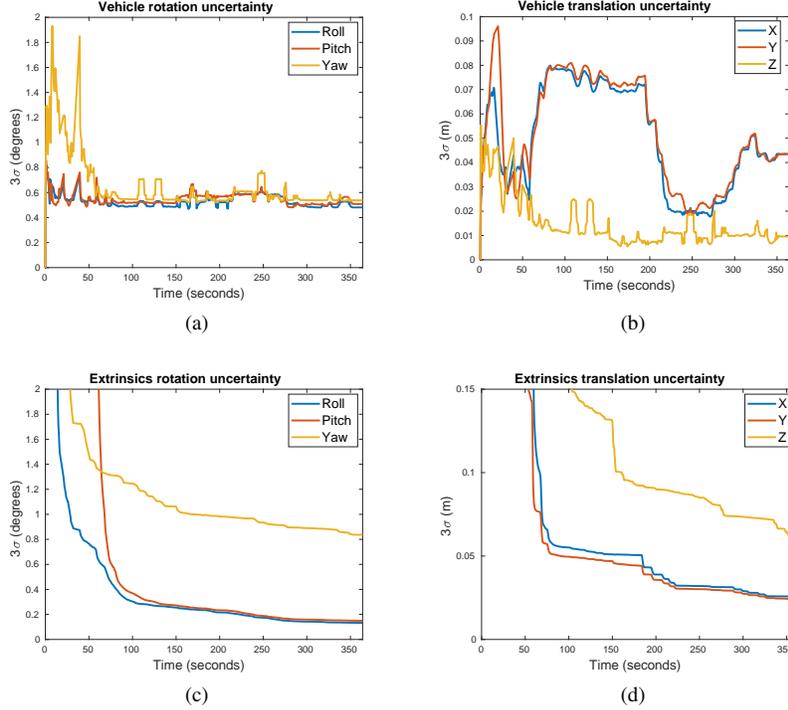


Figure 4: Three-sigma uncertainty bounds on: (a) the vehicle rotation (b) the vehicle translation (c) the extrinsics rotation and (d) the extrinsics translation. All quantities are evaluated on the same dataset using the proprietary vehicle odometry with no noise added.

3.6 Implementation

The proposed SLAM framework is implemented using the GTSAM library for optimization and on-manifold operations [4]. Since odometry and camera measurements arrive asynchronously, we consider each camera measurement as a single timestep in our framework and interpolate the odometry pose estimates using the camera measurement’s timestamp. The factor graph is optimized using the iSAM2 algorithm for efficient, real-time state estimation[11]. Analytical Jacobians are implemented for all factors in the nonlinear least-squares optimization except the XYH and ZPR factors, for which numerical Jacobians are computed.

4 Experimental results

4.1 Setup and evaluation metrics

The proposed SLAM system is evaluated using the HAUV platform described in Section 2 in an indoor 7m-diameter, 3m-deep test tank. We placed two of our custom-made boards on the tank floor, spanning an area of approximately 2x1 meters. Both the DVL and camera were fixed pointing downward. It is difficult to evaluate the accuracy of the localization of our algorithm without an underwater motion capture system. Previous works have used manual measurements [9] or ceiling-mounted vision systems [8] to obtain ground-truth trajectories of the AUV. However, our method is expected to be *more* accurate than manual measurements or a simple external vision system (except a highly

DVL-Camera Extrinsics

Trial	ψ (deg)	θ (deg)	ϕ (deg)	x (cm)	y (cm)	z (cm)
Accurate 1	-0.30	0.62	90.47	6.62	-23.67	11.75
Accurate 2	-0.28	0.52	90.06	5.94	-23.36	8.30
Accurate 3	-0.47	0.55	90.24	6.20	-23.35	13.08
Mean	-0.35	0.57	90.26	6.25	-23.46	11.04
Noisy 1	-0.301	0.62	90.30	6.03	-24.37	20.87
Noisy 2	-0.27	0.52	90.03	11.11	-26.50	23.11
Noisy 3	-0.46	0.55	90.56	8.51	-25.66	19.21
Mean	-0.34	0.56	90.30	8.55	-25.51	21.06
Manual	0	0	90	6	-23	8

Table 2: Estimated vehicle-camera extrinsics parameters from our real-world experiments. Extrinsics are shown in yaw, pitch, roll, and x , y , and z translation. The mean of each set of three datasets is bolded, and our manual measurement is shown for comparison.

DVL-Camera Extrinsics: Deviation from mean (σ)

Trial	ψ	θ	ϕ	x	y	z
Accurate 1	1.26	1.25	0.72	0.35	-0.21	0.31
Accurate 2	1.62	-0.95	-0.67	-0.30	0.10	-1.20
Accurate 3	-2.88	0.30	-0.05	-0.05	0.11	-0.89
Noisy 1	0.89	1.12	-0.01	-2.37	1.13	-0.08
Noisy 2	1.89	-1.26	-0.89	2.41	-0.98	0.85
Noisy 3	-2.79	0.13	0.89	-0.04	-0.15	-0.77

Table 3: Standard deviations of our experimental results from the sample mean, evaluated using the standard deviations derived from the corresponding marginal uncertainty in the overall factor-graph optimization. All estimates are within 3σ of the mean.

calibrated, multi-camera motion capture system) because it optimizes over high-quality odometry from the IMU and DVL as well as direct measurements of the AprilTag fiducials. Therefore, we validate this system statistically, by demonstrating with repeated trials that the resulting camera-vehicle extrinsics estimates have low variance and are consistent with the corresponding uncertainty in the factor-graph optimization.

Additionally, we evaluate the system using two different types of odometry: (1) the proprietary vehicle odometry that fuses IMU, depth sensor, and DVL measurements and (2) the proprietary vehicle odometry with random Gaussian noise added in the XYH directions (zero-mean, with standard deviation of 0.01 radians and 0.01 meters per frame). These are the degrees of vehicle motion that are not directly observable by the IMU or depth sensor. Therefore, this noisy odometry simulates an estimate that would be provided by just an IMU and depth sensor, without the DVL.

We recorded three datasets with which to perform SLAM and calibration. The vehicle was remotely operated and its motion consisted of translation along the x and y vehicle axes at various depths between 0 and 1.5 meters, and rotation about the z -axis (yaw rotation). This utilizes all controllable degrees of freedom of HAUV motion available, as the pitch and roll of the vehicle are not controllable by the thrusters. It is important to utilize all possible degrees of freedom of motion to provide as many constraints on the camera-vehicle extrinsics as possible.

4.2 Uncertainty

To demonstrate the bounded vehicle pose uncertainty and the convergence of the extrinsics estimate, we examine the marginal covariance of these variables at every step in the optimization. Figure 4 shows plots of the 3σ bounds of both the vehicle pose and the extrinsics for one of our experimental datasets, separating the values into rotation and translation uncertainty. The vehicle pose estimates are tightly bounded (0.5° rotation and 0.03m), except for times when the AprilTags either briefly go out of the field of view of the camera, or when the vehicle is too close to the tags to observe more than one or two in a single frame. The latter case is clearly visible in the vehicle translation uncertainty from 100 – 300 seconds, when the uncertainty in x and y translation rises as the vehicles dives close to the tank floor.

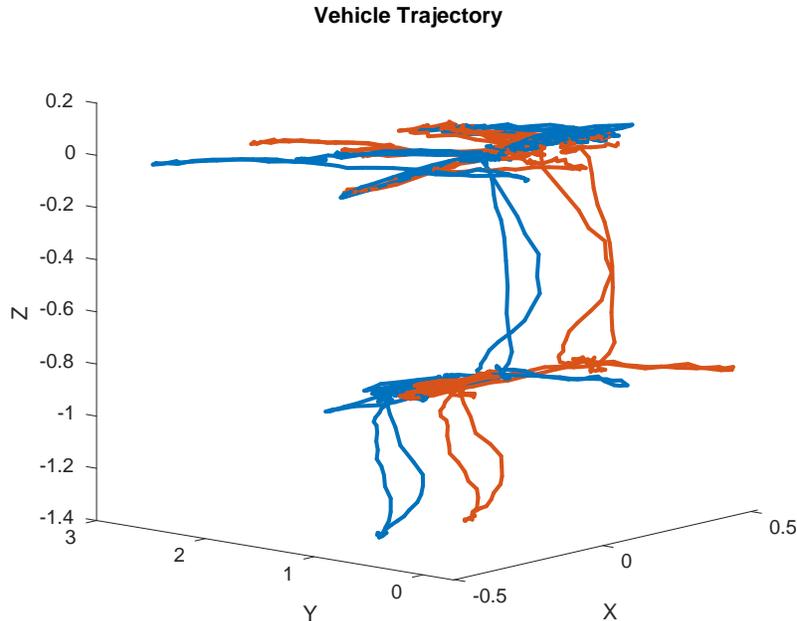


Figure 5: Sample dead-reckoning trajectory (red) and SLAM trajectory (blue) using noisy odometry measurements. The dead reckoning trajectory clearly drifts from the SLAM solution even over the course of a short, six-minute dataset.

4.3 Consistency

Table 2 shows the resulting extrinsics estimates for our six datasets. “Accurate” denotes the datasets using the vehicle odometry with no noise added. “Noisy” denotes the datasets using the vehicle odometry with noise added in the XYH directions. The mean extrinsics for both categories are also shown. In order to demonstrate the consistency of the extrinsics estimate, we examine the extrinsics estimate from these repeated trials with respect to the estimated uncertainty from the overall factor graph optimization. Since the extrinsics uncertainties at the end of the optimization are very similar across all three datasets, we arbitrarily use the uncertainty estimate from the first dataset. Table 3 shows the deviation of each dataset’s extrinsics estimates from the sample mean, normalized by the uncertainty. Most of the values lie within 2σ of the mean value, with all lying within 3σ . While three datasets is a small sample size, this confirms that our method is likely to provide a good upper-bound on the uncertainty of the extrinsics estimate.

Finally, we show in Figure 5 a comparison of the dead reckoning and SLAM trajectories, both utilizing the noisy odometry estimates. The dead reckoning estimate clearly drifts by tens of centimeters if not meters over the course of the six-minute dataset. The same trend may be seen using the accurate odometry estimates, but the difference between the trajectories is less pronounced.

5 Conclusion

We have presented a novel formulation of simultaneous underwater localization, mapping, and extrinsics calibration using a camera and one or more odometry sensors, such as an IMU and DVL. We utilize AprilTag fiducials to make our SLAM solution highly reconfigurable, inexpensive, and robust. The resulting extrinsics are consistent with the optimization’s uncertainty model, and very accurate in rotation.

One limitation of the proposed framework is the effect of timing synchronization errors on the system accuracy. This will present itself most significantly when the vehicle undergoes relatively high accelerations, as the velocity measurements made by the IMU and DVL may line up poorly with the camera measurements. A possible improvement would be to rigorously characterize the noise of the DVL measurements and develop a more accurate noise model. Additionally, it may be beneficial to explicitly model more vehicle dynamics, such as the mechanical slop in the roll cage on which the sensors are mounted, as in [14]. Finally, exploring recent extensions to AprilTag fiducials could make the system more robust to lighting changes that result in poor corner estimation [12].

Acknowledgment

This work was partially supported by the Office of Naval Research under awards N00014-16-1-2103 and N00014-16-1-2365

References

- [1] M. Agrawal, "A lie algebraic approach for consistent pose registration for general euclidean motion," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2006, pp. 1891–1897.
- [2] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *British Machine Vision Conference, BMVC 2013, Bristol, UK, September 9-13, 2013*, 2013.
- [3] F. Dellaert and M. Kaess, "Factor graphs for robot perception," *Foundations and Trends in Robotics*, vol. 6, no. 1-2, pp. 1–139, Aug. 2017, <http://dx.doi.org/10.1561/23000000043>.
- [4] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," GT RIM, Tech. Rep. GT-RIM-CP&R-2012-002, Sept 2012. [Online]. Available: <https://research.cc.gatech.edu/borg/sites/edu.borg/files/downloads/gtsam.pdf>
- [5] M. Fiala, "ARtag, a fiducial marker system using digital techniques," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Jun. 2005, pp. 590–596.
- [6] General Dynamics Mission Systems, "Bluefin HAUV," <https://gdmissionsystems.com/products/underwater-vehicles/bluefin-hauv>.
- [7] F. Hover, R. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. Leonard, "Advanced perception, navigation and planning for autonomous in-water ship hull inspection," *Intl. J. of Robotics Research*, vol. 31, no. 12, pp. 1445–1464, Oct. 2012.
- [8] J. Jung, Y. Lee, D. Kim, D. Lee, H. Myung, and H.-T. Choi, "AUV SLAM using forward/downward looking cameras and artificial landmarks," in *Underwater Technology (UT), 2017 IEEE*. IEEE, 2017, pp. 1–3.
- [9] J. Jung, J.-H. Li, H.-T. Choi, and H. Myung, "Localization of AUVs using visual information of underwater structures and artificial landmarks," *Intelligent Service Robotics*, vol. 10, no. 1, pp. 67–76, 2017.
- [10] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. Robotics*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.
- [11] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Intl. J. of Robotics Research, IJRR*, vol. 31, no. 2, pp. 217–236, Feb. 2012.
- [12] J. G. Mangelson, R. W. Wolcott, P. Ozog, and R. M. Eustice, "Robust visual fiducials for skin-to-skin relative ship pose estimation," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, 2016.
- [13] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2011.
- [14] P. Ozog, M. Johnson-Roberson, and R. M. Eustice, "Mapping underwater ship hulls using a model-assisted bundle adjustment framework," *Robotics and Autonomous Systems, Special Issue on Localization and Mapping in Challenging Environments*, vol. 87, pp. 329–347, 2017.
- [15] Qualisys AB, "Oqus Underwater," <https://www.qualisys.com/cameras/oqus-underwater/>.
- [16] P. Teixeira, M. Kaess, F. Hover, and J. Leonard, "Underwater inspection using sonar-based volumetric submaps," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, IROS*, Daejeon, Korea, Oct. 2016, pp. 4288–4295.
- [17] M. VanMiddlesworth, "Toward autonomous underwater mapping in partially structured 3d environments," Master's thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2014.
- [18] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg, "Pose tracking from natural features on mobile phones," in *IEEE and ACM Intl. Sym. on Mixed and Augmented Reality (ISMAR)*, 2008, pp. 125–134.