Ocean Engineering 280 (2023) 114675

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Development and experimental evaluation of visual-acoustic navigation for safe maneuvering of unmanned surface vehicles in harbor and waterway areas

Øystein Volden [a,*], David Cabecinhas [b], António Pascoal [b], Thor I. Fossen [a]

[a] Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway
[b] Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal

ARTICLE INFO

Keywords:
Velocity aiding
Landmark-based navigation
Error-state Kalman filter
Unmanned surface vehicles

ABSTRACT

Ship maneuvering in narrow harbor environments requires accurate navigation data to operate safely. Since Global Navigation Satellite Systems (GNSS) can be unreliable and inaccurate in such environments, other aiding techniques should be used for increased redundancy and reliability. To this end, we present a GNSS-free sensor suite consisting of visual, acoustic, and inertial sensors that can be used to navigate small vehicles in a harbor environment. The proposed method is aided by two navigation systems: a doppler velocity log (DVL) for low-drift velocity aiding and a visual fiducial system to estimate the drift-free position and attitude of a camera onboard the vehicle with respect to easily identifiable tags at the dockside. The visual and acoustic measurements are fused with inertial data using an error-state Kalman filter for robust state estimation. To benchmark the performance, we employ an unmanned surface vehicle equipped with a dual-antenna real-time kinematic GNSS receiver for accurate positioning and heading. The benefits of supplementing landmark-based navigation with acoustic measurements are demonstrated by conducting adverse scenarios that include noisy measurements and sensor dropouts in the harbor environment. As a result, the proposed visual-acoustic system can supplement satellite-based navigation systems for safety-critical harbor maneuvering.

1. Introduction

Lately, autonomy has gained increasing attention in the maritime sector (Relling et al., 2018). In the years to come, autonomous vehicles are expected to play an important role in industries such as shipping and public transportation due to reduced costs and improved safety (Kretschmann et al., 2017). Yet, several challenges must be overcome before fully autonomous vehicles are ready for the commercial market. In particular, autonomous vehicles must be equipped with highly robust and redundant navigation systems to operate reliably over time for widespread acceptance of commercial autonomy among authorities, classification societies, and the general public (Bolbot et al., 2020).

Ship maneuvering in constrained harbor and waterway areas requires accurate navigation data to operate reliably. Det Norske Veritas, an international ship classification company, requires autonomous ships to obtain 1 m absolute position accuracy with 95% probability for maneuvering in port areas (DNV, 2018). Differential Global Navigation Satellite Systems (GNSS) or real-time kinematic (RTK) GNSS can be used to exceed this level of accuracy. Unfortunately, satellite-based navigation signals are vulnerable to cyber–physical attacks such as spoofing and jamming (Carroll, 2003). Moreover, such signals suffer from multi-path effects in which the receiver also registers indirect signals reflected off the ground or other terrain obstacles (Prochniewicz and Grzymala, 2021). Satellite signals can also be blocked by structures such as buildings and bridges in urban areas, which make them unreliable, inaccurate, and sometimes unavailable (Hussain et al., 2021). For these reasons, researchers and ship classification societies require an independent navigation system that is more reliable in harbor and waterway areas (Androjna et al., 2020).

Researchers are increasingly focusing on visual-based localization systems because they are more robust, reliable, and cheaper than other sensor-based localization systems, e.g., ultrasonic or laser-based systems (Aqel et al., 2016). A significant part of the literature focuses on feature-based localization methods such as visual odometry (VO) or simultaneous localization and mapping (SLAM) systems because they are flexible, i.e., they require no additional infrastructure in the environment. While working well for certain indoor robotic applications, the performance of these techniques usually degrades in outdoor

* Corresponding author.
E-mail address: oystein.volden@ntnu.no (Ø. Volden).

https://doi.org/10.1016/j.oceaneng.2023.114675
Received 27 November 2022; Received in revised form 21 March 2023; Accepted 23 April 2023

environments due to textureless areas and challenging lighting conditions (Kalaitzakis et al., 2021). In contrast, vision-based localization methods aided by artificial tags (Kato and Billinghurst, 1999; Fiala, 2005; Garrido-Jurado et al., 2014; Olson, 2011; Wang and Olson, 2016) suffer less from these drawbacks. As such, these methods are preferred if accuracy, robustness, and speed are essential (Mondjar-Guerra et al., 2018). More importantly, visual fiducial systems do not induce drift since the camera-tag pose is estimated for each image independently of the previous one. Landmark-based localization can therefore function as a full-fledged absolute positioning alternative to GNSS if the target point is known in a global frame.

Unfortunately, fiducial tags are range-sensitive and limited by the camera's field of view (FOV). Moreover, the visual measurements can be inaccurate or completely absent if visibility is degraded. Hence, we also prefer to aid the inertial navigation system (INS) using other GNSS-free sensors. In this context, a doppler velocity log (DVL) can be used to complement the information obtained by the visual-inertial sensor suite (Romanovas et al., 2015; Fukuda et al., 2021; Ziebold et al., 2018). DVLs provide a flexible, energy-efficient solution in a small form factor that can produce accurate velocity measurements as long as the acoustic beams reach the sea bottom while operating in *bottom-lock* mode. Bottom-lock mode refers to situations where at least three beam measurements have a line of sight to the sea floor (Rudolph and Wilson, 2012). Although small velocity errors are integrated up and will, over time, result in less accurate position estimates, a DVL is very useful for maintaining a low-drift estimate until absolute corrections are available. To bound the inertial navigation error, terrain-based navigation relying on bathymetric measurements can also be used. However, this requires an a priori bathymetry reference map which can be resource-intensive to produce (Salavasidis et al., 2019). As such, we will focus on acoustic-based localization, which does not require a reference map and is, therefore, more flexible.

A Kalman filter (Farrell, 2008) is usually used to fuse high-rate inertial measurement unit (IMU) measurements with low-rate aiding sensor measurements, e.g., camera and DVL measurements, similar to standard GNSS + INS solutions (Farell and Barth, 1998). To capture the nonlinearities in the process and measurement model, nonlinear formulations of the Kalman filter, e.g., the *Extended Kalman filter* (EKF) and the *unscented Kalman filter* (UKF) (Julier and Uhlmann, 2004), are often used. While the UKF is beneficial for highly nonlinear systems, it is computationally expensive compared to the EKF (Biswas et al., 2016). Therefore, the EKF is usually considered for real-time state estimation of navigation signals. The accuracy of the EKF is, however, highly dependent on how well the nonlinearities of the models can be captured by linearizations about operating points. For example, if the vehicle dynamics exhibit strong nonlinear behavior, the local linearization of the states in the EKF will not represent a sufficiently accurate approximation. Therefore, the *error-state Kalman filter* (ESKF) was invented to improve upon the linear approximation. By estimating the error state instead of the true state, the ESKF allows for better linearization since the error state tends to be approximately linear, i.e., higher-order products become negligible (Sola, 2017). Another benefit is dead-reckoning performance due to more accurate bias estimation (Fossen, 2021, p. 471). Furthermore, a four-component unit quaternion is advantageous when representing attitude. As such, the multiplicative extended Kalman filter (MEKF) formulation is preferred: an error-state EKF where the attitude is parametrized using a four-dimensional unit quaternion (Markley and Crassidis, 2014). However, the unit quaternion error is parametrized using a three-parameter attitude representation. This is advantageous since the three-parameter representation avoids singularities due to small attitude errors (Crassidis et al., 2007). After estimating the error state, the latter is injected into the nominal state to predict the true state using the four-dimensional unit quaternion representation instead of the minimal three-parameter representation. Finally, a reset strategy (feedback filter) is employed in which the error state is set to zero to prevent the state estimates from accumulating for long-endurance missions (Fossen, 2021, p. 472).

In this paper, we fuse drift-free (visual) and low-drift (acoustic) measurements with inertial data using a MEKF. As such, we obtain a nonsingular attitude representation with sufficient linearization properties, resulting in robust and accurate state estimates. We also show how the acoustic measurements can supplement the camera solution to extend the scope in which the proposed filter can aid the vehicle. Hence, we contribute to the development of an independent, GNSS-free navigation system for increased accuracy and redundancy in safety-critical maritime operations.

1.1. Related work

Since feature-based localization methods such as VO or SLAM systems are usually prone to fail totally or partially in outdoor environments (Kalaitzakis et al., 2021), researchers are looking into supplementary components that can improve localization. One of these is fiducial markers: artificial landmarks that are designed to be very distinct and, therefore, easily identifiable in outdoor environments. As such, the following literature review comprises work on vision-based navigation aided by artificial landmarks with a particular focus on outdoor maritime environments.

Since fiducial tags function in GNSS-denied environments, they have become a popular choice to aid vehicles in underwater operations. For example, Myint et al. (2018) and Hsu et al. (2022) combined stereo vision with customized 3-D markers to develop a landmark-based docking system for autonomous underwater vehicles. Further, Chen et al. (2021) used a collection of Apriltags to estimate the pose of a remotely operated vehicle (ROV) underwater, and Trslic et al. (2020) tested a vision-based docking system for ROVs in the North Atlantic Ocean by combining monocular vision with customized light beacons. Landmark-based aiding techniques have also been investigated overwater. For example, Volden et al. (2021) developed a hybrid data-driven and model-based scheme to estimate the position of an unmanned surface vehicle (USV) aided by ArUco markers. Also, in the air, Malyuta et al. (2020) designed an unmanned aerial vehicle system assisted by AprilTags for precise landing and automatic charging.

A significant drawback of the aforementioned works is that the fiducial tags are range-constrained and limited by the camera's FOV and resolution. As such, vision-based navigation aided by artificial tags has practical limitations if the vehicle is operating over larger areas. In maritime environments, however, acoustic-based localization can be used to supplement landmark-based localization methods. For example, Xu et al. (2021) and Vargas et al. (2021) fused visual and acoustic measurements using a stereo camera and a DVL in a graph SLAM framework for underwater localization and mapping. Although they utilized acoustic measurements for increased robustness, feature-based localization underwater is difficult due to poor visibility, lack of light, and insufficient texture in the scene (Vargas et al., 2021). However, vision-based solutions suffer less from these difficulties overwater, especially those assisted by artificial tags. Therefore, for visual-acoustic sensor suites onboard surface vehicles, we suggest the use of visual measurements above the water surface combined with underwater acoustic measurements for enhanced navigation capability. Contrary to our work, the combination of these aiding techniques has rarely been demonstrated experimentally onboard a USV.

1.2. Main contribution

This paper presents a case study that showcases the effective utilization of a well-known multi-modal sensor fusion scheme driven by visual, acoustic, and inertial measurements to accurately estimate the full state of a vehicle operating in a harbor environment. We have developed a sensor fusion algorithm that can be used as an independent and reliable alternative to satellite-based navigation, thus increasing

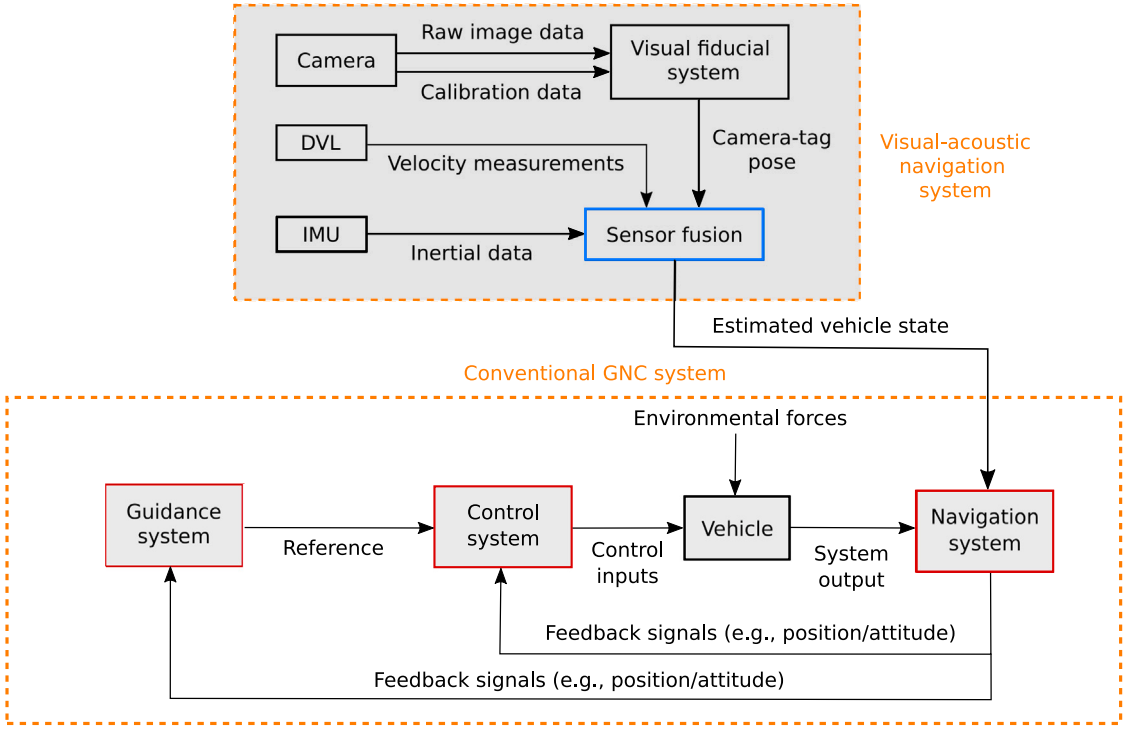


Fig. 1. An overview of the proposed visual-acoustic navigation system and its relation to the conventional guidance, navigation, and control (GNC) system.

the redundancy and accuracy for safety-critical operations in harbor and waterway areas. To attest to the accuracy of the proposed solution, we use a synchronized experimental dataset with visual, acoustic, and inertial data and highly accurate RTK GNSS measurements for benchmarking. Results show that our custom ESKF implementation, aided by camera-tag pose and DVL measurements, performs well for different scenarios in the harbor. We argue that our study is unique, as we provide a complete description of how acoustic velocity measurements supplement landmark-aided navigation in a sensor fusion scheme to enhance GNSS-free navigation. We have developed source code to run the algorithms in the Robot Operating System (ROS) (ROS, 2022) available in a public GitHub repository (Volden, 2022). The filter implementation is based on the MEKF algorithm described in the Marine Systems Simulator toolbox (Fossen and Perez, 2004) but interfaced and customized to our sensor suite.

1.3. Outline

The paper is structured as follows. Section 2 describes the proposed multi-modal sensor fusion scheme, including a DVL for velocity aiding, a visual fiducial system for range-constrained position and attitude aiding, and an ESKF to fuse the visual, acoustic, and inertial measurements. In Section 3, the experimental setup is described, followed by an experimental description. Then, in Section 4, we present and discuss the experimental results. Finally, in Section 5 we present the conclusion and discuss relevant issues that warrant further research.

2. Methodology

2.1. System overview

The proposed method consists of three main components: a DVL for low-drift velocity aiding, a visual fiducial system to measure the relative camera-tag pose, and a sensor fusion scheme to estimate the full state of the vehicle driven by visual, acoustic, and inertial measurements. Given pre-determined coordinate transformations in the sensor suite, the proposed visual-acoustic navigation system produces the estimated state of the vehicle in the North-East-Down (NED) frame based on an ESKF and strapdown navigation equations. These estimates are then used by the conventional navigation system, as shown in Fig. 1. The main motivation is to develop a resilient navigation system for USVs in harbor and waterway areas by providing an independent, GNSS-free navigation solution that is highly accurate in fully or partly GNSS-denied environments. For example, the filter can immediately use the IMU/DVL solution to minimize the vehicle's navigation errors under a GNSS dropout. However, due to small velocity errors induced by the DVL, the integrated position estimates will slowly drift, and absolute corrections are required for long-endurance applications. Since GNSS dropouts can continue for an unpredictable period, it is desirable to have a backup sensor. For this purpose, we use artificial tags at the dockside combined with a camera to estimate the relative camera-tag pose, which provides drift-free measurements as input to the sensor fusion scheme. As such, we provide a flexible IMU/DVL solution for short-endurance operation supported by a camera solution for long-endurance operation, given that the vehicle is closer than a certain distance to the landmarks.

2.2. Doppler velocity log

A DVL uses hydro-acoustic beams to measure the velocity of the vehicle with respect to the sea bottom and the distance to the latter. The four-transducer configuration with beams pointing forward, backward, left, and right relative to the vehicle is the most common solution, also referred to as the *Janus configuration*. The DVL is mounted on the surface vehicle such that each transducer is pointed towards the sea floor.

2.2.1. Velocity estimation

Initially, an acoustic signal is sent from each of the four transducers. Then, each signal will reflect off the sea bottom and return to the transducers. The doppler measurement is then performed: a frequency shift between the transmitted and returned signal for each transducer. By using the frequency shift, velocities along the DVL-fixed transducer axes are computed. In addition, the time of flight is measured to

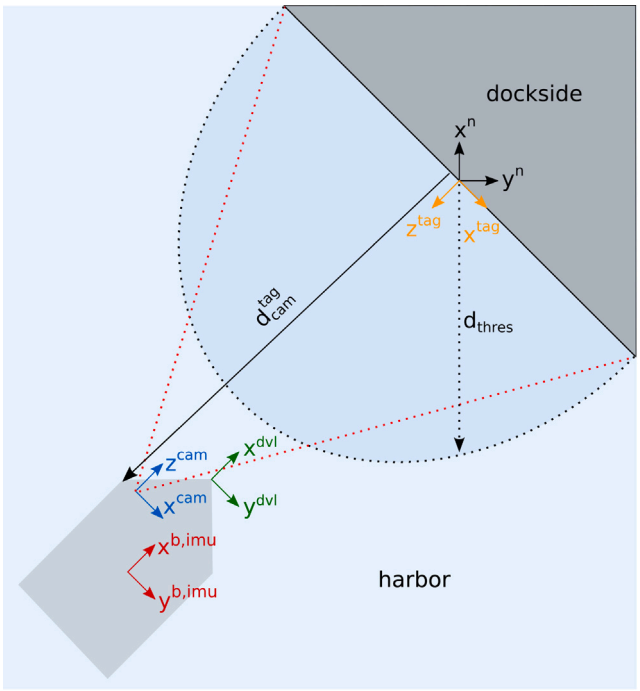


Fig. 2. An overview of the coordinate systems onboard the vehicle and at the dockside. The figure also shows the measured camera-tag Euclidean distance $d_{\text{cam}}^{\text{tag}}$ and the associated threshold d_{thres}.

estimate the distance between each transducer and the seabed. Then, the onboard attitude and heading reference system (AHRS)/IMU reads its triaxial gyroscope, accelerometer, and compass sensors to determine orientation. Finally, the measured velocity for each transducer and AHRS orientation are fused using a Kalman filter to estimate the velocity of the DVL with respect to the sea bottom. The DVL measurement is usually accepted if the DVL operates in *bottom lock* mode.

2.3. Visual fiducial system

Fiducial markers are artificial landmarks of known size and shape that feature a specific pattern that is used to detect and localize them. In our work, we employ a ROS wrapper (Malyuta and Wolfgang, 2022) of *AprilTag* (Olson, 2011): a black-and-white square fiducial marker system. The visual fiducial system consists of three main components: a detector, a coding system, and a pose estimation algorithm.

2.3.1. Detector, coding system and pose estimation

Initially, the detector tries to find four-sided regions, referred to as a *quad*. By intersecting the line segments forming the quad, the detector obtains the four corners of the tag. Once detected, a digital coding system assigns a unique ID for each marker based on the number of bits to create the marker's codeword, that is, its inner pattern. The available codewords are carefully selected to be robust and error-correcting, allowing them to be used for longer ranges and non-optimal visibility conditions. We use tags with 6×6 tag size, i.e., 36-bit codeword, to obtain a good tradeoff between speed and accuracy. Then, when the tag is detected and uniquely classified, the image is undistorted using radial and tangential distortion coefficients before image-to-point correspondences are extracted from the tag corners. These correspondences are then used by the Point-n-Perspective solver, along with the camera calibration matrix and the physical size of the tag, to estimate the camera-tag pose. The physical size of the tag is used to overcome the scale ambiguity such that the absolute pose can be reconstructed from single-view geometry. We refer to Olson (2011) for a more detailed description of the AprilTag system and the reconstruction of the pose from the tag corners.

2.3.2. Acceptance criteria

Fiducial tags are range-sensitive and limited by the camera's FOV. Hence, the camera onboard the vehicle needs to be closer than a certain distance from the tags and point towards them to obtain camera-tag pose measurements. However, since the measurements can be noisy and inaccurate at large distances, we want to specify a distance threshold in which the camera solution produces sufficiently accurate measurements. This can be achieved by comparing AprilTag position and heading measurements with ground truth data offline, e.g., from dual-antenna RTK GNSS. By specifying an acceptable threshold in error between AprilTag and ground truth, we can examine the corresponding camera-tag distance at which the AprilTag system produces a lower error than the pre-specified threshold. We denote this 2-D Euclidean distance threshold as d_{thres}, which is illustrated in Fig. 2. Hence, the camera measurement is accepted if the measured 2-D Euclidean distance $d_{\text{cam}}^{\text{tag}}$ is lower than d_{thres}.

Unfortunately, a single-tag configuration with n = 4 point correspondences is prone to rotational ambiguity when in weak-perspective conditions, that is, when imaging small marker planes or planes at distances considerably larger than the camera's focal length (Collins and Bartoli, 2014). This is because four correspondences per tag are used, leading to a lack of redundancy such that the homography overfits. For $n > 4$, there is redundancy, leading to considerably lower error (Collins and Bartoli, 2014). As a result, we require that two or more tags, i.e., $n \geq 8$, can be detected simultaneously in an image to use the pose measurement from the visual fiducial system. A tag rejection algorithm based on the range constraint and the number of detected tags required is summarized in Algorithm 1.

Algorithm 1 Tag rejection algorithm

n_{tag}: Number of detected tags
$[x_{\text{cam}}^{\text{tag}}, z_{\text{cam}}^{\text{tag}}]$: Camera-tag 2-D position measurement
d_{thres}: 2-D Euclidean distance threshold
B_{accept}: Boolean acceptance flag

Input: $(n_{\text{tag}}, x_{\text{cam}}^{\text{tag}}, z_{\text{cam}}^{\text{tag}})$
Output: B_{accept}

for each image **do**
  if tag detected **then**
    $d_{\text{cam}}^{\text{tag}} \leftarrow \sqrt{(x_{\text{cam}}^{\text{tag}})^2 + (z_{\text{cam}}^{\text{tag}})^2}$
    if $(d_{\text{cam}}^{\text{tag}} < d_{\text{thres}})$ and $n_{\text{tag}} \geq 2$ **then**
      $B_{\text{accept}} \leftarrow$ true (accept measurement)
    else
      $B_{\text{accept}} \leftarrow$ false (reject measurement)
    end if
    return B_{accept}
  end if
end for

2.4. Coordinate transformations

For strapdown INS, it is common to express the aiding sensors in the NED frame, with IMU measurements in the body frame. As such, we transform the camera measurements, the DVL measurements, and the GNSS measurements to NED coordinates. We emphasize that the IMU is located at the center of the vehicle with orientation aligned with the body frame. Hence, no lever-arm compensation is necessary. Also, note that the USV is symmetric about the planes orthogonal to the main axes and is relatively small, i.e., 2 m long and 1 m wide. Therefore, we assume that the center of the vehicle is close to the center of mass.

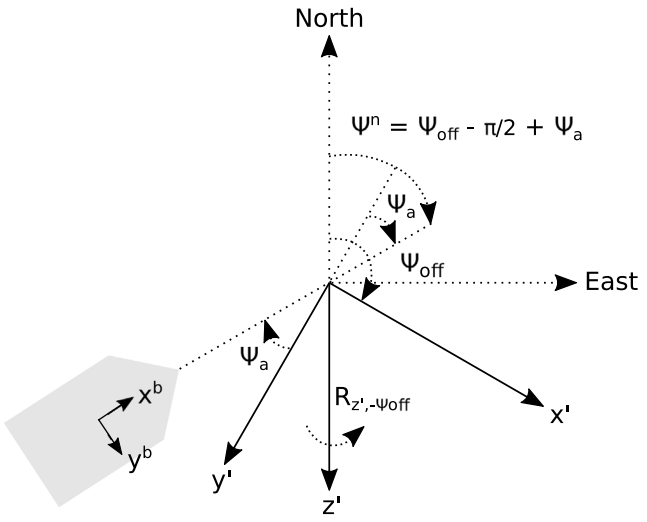


Fig. 3. The heading of the vehicle in the NED frame can be computed from the angle of the vehicle relative to the axis pointing out of the tag (ψ_a) and the fixed yaw offset between the true north and the x-axis of the tag frame (ψ_{off}).

2.4.1. Camera: tag to NED

Let $\boldsymbol{T}_{\text{cam}}^{\text{tag}}$ be the transformation matrix containing the rotation and translation of the camera in the tag frame, computed by the visual fiducial system. Furthermore, let $\boldsymbol{T}_{b}^{\text{cam}}$ be the transformation matrix that contains the fixed in-vehicle translation from the camera to the center of the vehicle in the camera frame. Finally, let $\boldsymbol{T}_{\text{tag}}^{n}$ be the transformation matrix that contains the rotation matrix to rotate from the tag frame to the NED frame. For simplicity, the NED and tag frame have identical origins, both located in the middle of the reference tag (see Fig. 2). By concatenating the transformations, we get

$$\begin{aligned}\boldsymbol{T}_{b}^{n} &= \begin{bmatrix}\boldsymbol{R}_{b}^{n} & \boldsymbol{t}_{b}^{n}\\ 0 & 1\end{bmatrix} = \boldsymbol{T}_{\text{tag}}^{n}\boldsymbol{T}_{\text{cam}}^{\text{tag}}\boldsymbol{T}_{b}^{\text{cam}}\\ &= \begin{bmatrix}\boldsymbol{R}_{z',-\psi_{\text{off}}}\boldsymbol{R}_{x,\pi/2} & \boldsymbol{0}_{3\times1}\\ 0 & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{R}_{\text{cam}}^{\text{tag}} & \boldsymbol{t}_{\text{cam}}^{\text{tag}}\\ 0 & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{I}_{3} & \boldsymbol{t}_{b}^{\text{cam}}\\ 0 & 1\end{bmatrix},\end{aligned} \tag{1}$$

where $\boldsymbol{R}_{x,\pi/2} \in \text{SO}(3)$ rotates the x-axis of the tag frame by $\pi/2$, and $\boldsymbol{R}_{z',-\psi_{\text{off}}} \in \text{SO}(3)$ rotates the z-axis of the subsequent frame by $-\psi_{\text{off}}$. The last rotation, $\boldsymbol{R}_{z',-\psi_{\text{off}}}$, finalizes the transformation to the NED frame and is shown in Fig. 3. Hence, $\boldsymbol{T}_{\text{tag}}^{n}$ and $\boldsymbol{T}_{\text{b}}^{\text{cam}}$ are known. Further, $\boldsymbol{T}_{\text{cam}}^{\text{tag}}$ can be recovered since AprilTag computes the translation $\boldsymbol{t}_{\text{cam}}^{\text{tag}}$ directly, and the unit quaternion $\boldsymbol{q}_{\text{cam}}^{\text{tag}}$ can be mapped to the rotation matrix $\boldsymbol{R}_{\text{cam}}^{\text{tag}} \in \text{SO}(3)$ using the following formula (Fossen, 2021, p. 34):

$$\boldsymbol{R}_{\text{cam}}^{\text{tag}} = \boldsymbol{R}(\boldsymbol{q}_{\text{cam}}^{\text{tag}}) = \boldsymbol{I}_{3} + 2\eta\boldsymbol{S}(\epsilon) + 2\eta\boldsymbol{S}^{2}(\epsilon), \tag{2}$$

where η and ϵ are the real and imaginary components of the unit quaternion $\boldsymbol{q}_{\text{cam}}^{\text{tag}}$, and $\boldsymbol{S} \in \text{SS}(3)$ is a 3×3 matrix in the set of skew-symmetric matrices of order 3. Finally, the attitude of the vehicle in the NED frame can be recovered by extracting $\boldsymbol{R}_{b}^{n} \in \text{SO}(3)$ from (1) before transforming it into the unit quaternion $\boldsymbol{q}_{b}^{n}$ (Fossen, 2021, p. 34). Also, the center position of the vehicle in the NED frame is computed as

$$\boldsymbol{p}_{nb}^{n} = \boldsymbol{R}_{b}^{n}\boldsymbol{p}_{nb}^{b}, \tag{3}$$

where $\boldsymbol{p}_{nb}^{b}$ is the relative position between the center of the vehicle and the NED origin expressed in the body frame.

2.4.2. DVL: DVL to NED

Since the DVL is placed in front of the USV, thus measuring the velocity of a point that is different from the origin of the body reference axis, compensation due to angular velocity is necessary. To achieve this, we use a fixed lever-arm vector $\boldsymbol{l}^{b} = [x_l, y_l, z_l]^{T}$ between the body frame and the DVL expressed in the body frame to transform DVL-fixed velocity to body-fixed velocity. Further, no orientation offsets between the DVL and the body frame exist. As such, we can express the body-fixed linear velocity as

$$\boldsymbol{v}_{\text{dvl}}^{b} = \boldsymbol{v}^{\text{dvl}} + \boldsymbol{S}(\boldsymbol{\omega}_{\text{nb}}^{b})\boldsymbol{l}^{b}, \tag{4}$$

where $\boldsymbol{v}^{\text{dvl}}$ is the DVL-fixed linear velocity measured by the DVL and $\boldsymbol{S}(\boldsymbol{\omega}_{\text{nb}}^{b})$ is the skew-symmetric matrix of the body-fixed angular velocity vector $\boldsymbol{\omega}_{\text{nb}}^{b}$. Finally, the linear velocity can be expressed in NED as

$$\boldsymbol{v}_{\text{dvl}}^{n} = \boldsymbol{R}(\boldsymbol{q}_{b}^{n})\boldsymbol{v}_{\text{dvl}}^{b}, \tag{5}$$

where $\boldsymbol{R}(\boldsymbol{q}_{b}^{n}) \in \text{SO}(3)$ is the rotation matrix from body to NED.

2.4.3. GNSS: WGS-84 to NED

To obtain the North-East positions (x^n, y^n) from longitude–latitude (l, μ), we need to compute the Earth radius of curvature in the prime vertical R_N and the radius of curvature in the meridian R_M. They are defined by (Farrell, 2008)

$$R_N = \frac{r_e}{\sqrt{1 - e^2\sin^2(\mu_0)}} \tag{6}$$

$$R_M = R_N\frac{1 - e^2}{1 - e^2\sin^2(\mu_0)}, \tag{7}$$

where $r_e = 6\,378\,137$ m is the semi-minor axis (Equatorial radius), and $e = 0.0818$ is the Earth eccentricity. Given the measured longitude–latitude origin (l_0, μ_0), located in the middle of the reference tag, the North-East positions (x^n, y^n) can then be computed from (l, μ) by

$$x^{n} = \frac{\Delta\mu}{\text{atan2}(1, R_M)} \tag{8}$$

$$y^{n} = \frac{\Delta l}{\text{atan2}(1, \cos(\mu_0))}, \tag{9}$$

where $(\Delta l, \Delta\mu) := (l - l_0, \mu - \mu_0)$. We neglect the height transformation since the vehicle operates in 2-D, that is, on the surface of the geoid. Note that GNSS position and velocity are expressed in the NED frame relative to the body frame. We emphasize that the GNSS attitude is also expressed in the NED frame.

2.5. The error-state Kalman filter

The MEKF is an *indirect* filter technique in which the Kalman filter is formulated as an error-state filter. The error state $\delta\boldsymbol{x}$ includes position, velocity, attitude, and IMU bias errors. Since the IMU is strapped to the vehicle, the INS state $\boldsymbol{x}_{\text{ins}}$ is based on the integration of *strapdown navigation equations* to express the vehicle's motion. These equations are driven by IMU specific force and gyroscope measurements. In contrast, the model-based navigation filter uses a mathematical model driven by control inputs to generate linear acceleration and angular velocity. As such, it must compensate for wind and water currents to aid with respect to the inertial frame. This is, however, avoided in the strapdown INS since the IMU will pick up any motions acting on the vehicle. Further, the IMU measurements are integrated to obtain position and attitude, which will cause drift due to sensor biases, misalignments, and temperature variations (Fossen, 2021, p. 476). To compensate for the sensor biases, the filter is implemented as a feedback filter where the error estimates are used to update the INS estimates directly, as seen in Fig. 4. Hence, the estimated error state $\delta\hat{\boldsymbol{x}}$ is estimated by the MEKF before it is injected into the estimated INS state $\hat{\boldsymbol{x}}_{\text{ins}}$. In particular, a reset strategy is used by setting the error state to zero after each correction from drift-free measurements. As such, we ensure that $\hat{\boldsymbol{x}}_{\text{ins}} \rightarrow \boldsymbol{x}$ when the error-state estimate is fed back to the strapdown navigation equations, thus preventing the INS errors from growing due to improved bias estimation through drift-free measurements.

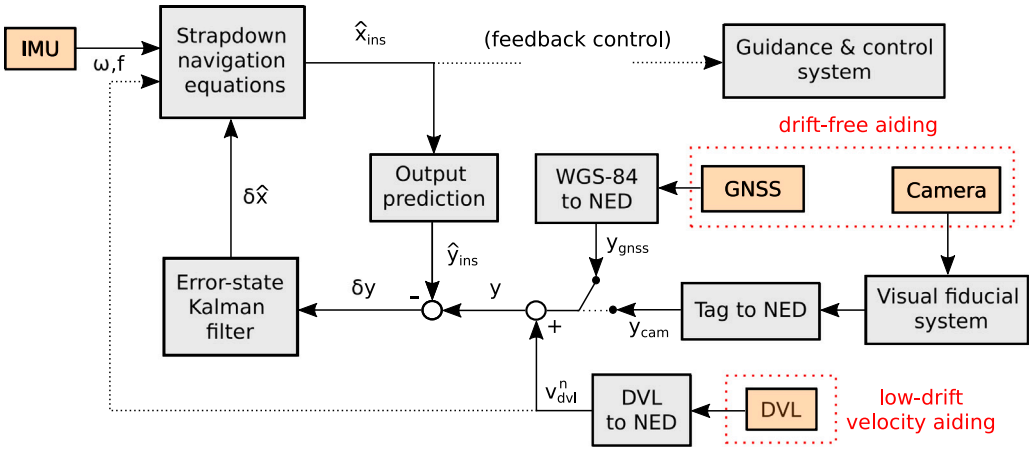


Fig. 4. An overview of the error-state filter aided by drift-free and low-drift sensor measurements. The filter receives either GNSS or camera measurements, depending on sensor availability.

2.5.1. Error-state dynamics

The core idea of the MEKF is that the unit quaternion error

$$\delta \boldsymbol{q}_b^n = \begin{bmatrix} \delta\eta \\ \delta\boldsymbol{\epsilon} \end{bmatrix} \tag{10}$$

can be parametrized using a three-parameter attitude representation where $\delta\eta$ and $\delta\boldsymbol{\epsilon}$ are the real and imaginary components of the unit quaternion error $\delta\boldsymbol{q}_b^n$. We will use the Gibbs vector $\boldsymbol{a}_g = [g_1, g_2, g_3]^T$ scaled by a factor of two for the mapping between the three-parameter attitude representation and the unit quaternion error:

$$\delta \boldsymbol{a}_g = \frac{\delta\boldsymbol{\epsilon}}{\delta\eta}, \ \delta\boldsymbol{a} := 2\delta\boldsymbol{a}_g, \ \delta\boldsymbol{q}_b^n = \frac{1}{\sqrt{4 + \delta\boldsymbol{a}^T\delta\boldsymbol{a}}} \begin{bmatrix} 2 \\ \delta\boldsymbol{a} \end{bmatrix}. \tag{11}$$

The scaled Gibbs vector is used to obtain physical quantities in the Kalman filter covariance matrix. As such, the covariance matrix estimates will be expressed in radians squared, which is equivalent to angle errors using a first-order approximation (Fossen, 2021, p. 481). With the three-parameter attitude error introduced using the Gibbs vector representation, we define the error state as

$$\delta\boldsymbol{x} = [(\delta\boldsymbol{p}_{nb}^n)^T, (\delta\boldsymbol{v}_{nb}^n)^T, (\delta\boldsymbol{b}_{\text{acc}}^b)^T, \delta\boldsymbol{a}^T, (\delta\boldsymbol{b}_{\text{gyro}}^b)^T]^T, \tag{12}$$

where $\delta\boldsymbol{b}_{acc}^b$ and $\delta\boldsymbol{b}_{gyro}^b$ denote the accelerometer and gyroscope bias error, respectively. Note that the state space is reduced from a 16 to a 15-state system since the $\delta\boldsymbol{a}$ vector replaces the unit quaternion error $\delta\boldsymbol{q}_b^n$. Further, the differential equations describing the error-state dynamics must be linearized such that they fit into the discrete-time system matrices. Hence, the error-state dynamics can be approximated by the first-order linear differential equations:

$$\delta\dot{\boldsymbol{p}}_{nb}^n = \delta\boldsymbol{v}_{nb}^n \tag{13}$$

$$\begin{aligned}\delta\dot{\boldsymbol{v}}_{nb}^n \approx &-\boldsymbol{R}(\hat{\boldsymbol{q}}_{\text{ins}})\boldsymbol{S}(\boldsymbol{f}_{\text{nb}}^b - \hat{\boldsymbol{b}}_{\text{acc}}^b)\delta\boldsymbol{a} \\ &- \boldsymbol{R}(\hat{\boldsymbol{q}}_{\text{ins}})(\delta\boldsymbol{b}_{\text{acc}}^b + \boldsymbol{w}_{\text{acc}}^b)\end{aligned} \tag{14}$$

$$\delta\dot{\boldsymbol{b}}_{\text{acc}}^b = -\frac{1}{T_{\text{acc}}}\delta\boldsymbol{b}_{\text{acc}}^b + \boldsymbol{w}_{b,\text{acc}}^b \tag{15}$$

$$\delta\dot{\boldsymbol{a}} \approx -\boldsymbol{S}(\boldsymbol{\omega}_{\text{nb}}^b - \hat{\boldsymbol{b}}_{\text{gyro}}^b)\delta\boldsymbol{a} - \delta\boldsymbol{b}_{\text{gyro}}^b - \boldsymbol{w}_{\text{gyro}}^b \tag{16}$$

$$\delta\dot{\boldsymbol{b}}_{\text{gyro}}^b = -\frac{1}{T_{\text{gyro}}}\delta\boldsymbol{b}_{\text{gyro}}^b + \boldsymbol{w}_{b,\text{gyro}}^b, \tag{17}$$

where $\boldsymbol{f}_{\text{nb}}^b$ is the specific force vector, and $\boldsymbol{\omega}_{\text{nb}}^b$ is the angular velocity vector, both expressed in the body frame. Further, T_{gyro} and T_{acc} are time constants that ensure that the bias errors go exponentially to zero during dead reckoning. They are usually found by trial and error but can also be obtained by computing the Allan variance (Rubiola, 2008) of the gyrocope and the specific force sensor and using this information to fit the bias models. The additive zero-mean Gaussian white noise terms $\boldsymbol{w}_{\text{acc}}^b$, $\boldsymbol{w}_{\text{gyro}}^b$, $\boldsymbol{w}_{b,\text{acc}}^b$, and $\boldsymbol{w}_{b,\text{gyro}}^b$, are used to model the measurement and bias noise, respectively. Also, note that $\hat{\boldsymbol{q}}_{\text{ins}} \equiv \hat{\boldsymbol{q}}_b^n$. We refer to Fossen (2021, p. 481–483) for the details regarding the derivation of the differential equations describing the error state dynamics.

2.5.2. Kalman filter measurements

For positioning purposes, GNSS is usually the primary sensor for aiding surface vehicles. We will, however, replace GNSS with the landmark-based solution and use it if the criteria in Algorithm 1 are satisfied. For velocity measurements, we will use the DVL. Note that the visual fiducial system does not measure velocity directly. It only measures the position and attitude between the camera and the reference object for each image, independently of the previous ones. The aiding measurements are expressed in the NED frame after using the coordinate transformations described in Section 2.4. Hence, the error-measurement equations for position and velocity can be described by

$$\delta\boldsymbol{y}_p = (\boldsymbol{p}_{nb}^n + \boldsymbol{\epsilon}_p) - \hat{\boldsymbol{p}}_{nb}^n = \delta\boldsymbol{p}_{nb}^n + \boldsymbol{\epsilon}_p, \tag{18}$$

$$\delta\boldsymbol{y}_v = (\boldsymbol{v}_{nb}^n + \boldsymbol{\epsilon}_p) - \hat{\boldsymbol{v}}_{nb}^n = \delta\boldsymbol{v}_{nb}^n + \boldsymbol{\epsilon}_p, \tag{19}$$

where $\boldsymbol{p}_{nb}^n$ and $\boldsymbol{v}_{nb}^n$ are the measured position and velocity in the NED frame, $\hat{\boldsymbol{p}}_{nb}^n$ and $\hat{\boldsymbol{v}}_{nb}^n$ are the estimated position and velocity in the NED frame, and $\boldsymbol{\epsilon}_p$ is assumed to be Gaussian white measurement noise.

In order to estimate the unit quaternion for attitude determination, a heading reference is needed to guarantee observability. Commercial ships usually combine the gravity vector with a high-quality gyrocompass for this purpose (Fossen, 2021, p. 484). We will, however, replace the gyrocompass with the camera solution and use it if the criteria in Algorithm 1 are satisfied. Regarding the gravity vector, we start by defining the normalized specific force vector as

$$\boldsymbol{v}_1^b := -\frac{\boldsymbol{f}^b}{g(\mu)}, \tag{20}$$

where the World Geodetic System 1984 (WGS-84) ellipsoidal gravity formula (Mularie, 2000) is used to compute $g(\mu)$ based on latitude μ, and $\boldsymbol{f}^b$ is the unbiased specific force vector expressed in the body frame. Then, the estimated normalized specific force vector is

$$\hat{\boldsymbol{v}}_1^b = \boldsymbol{R}^T(\hat{\boldsymbol{q}}_b^n)\boldsymbol{v}_{01}^n, \tag{21}$$

where $\boldsymbol{v}_{01}^n = [0, 0, 1]^T$ is chosen as the *gravity reference vector*, pointing downwards with respect to the NED frame, and $\boldsymbol{R}^T(\hat{\boldsymbol{q}}_b^n) \in \text{SO}(3)$ represents the estimated unit quaternion rotation matrix from the NED frame to the body frame. The following error-measurement equation,

describing the error gravity reference vector $\delta \boldsymbol{v}_1$, can be approximated as

$$\begin{aligned}\delta \boldsymbol{v}_1 &= (\boldsymbol{v}_1^b + \boldsymbol{\epsilon}_1) - \hat{\boldsymbol{v}}_1^b = (\boldsymbol{v}_1^b + \boldsymbol{\epsilon}_1) - \boldsymbol{R}^T(\hat{\boldsymbol{q}}_b^n)\boldsymbol{v}_{01}^n \\ &\approx \boldsymbol{S}(\boldsymbol{R}^T(\hat{\boldsymbol{q}}_{\text{ins}})\boldsymbol{v}_{01}^n)\delta\boldsymbol{a} + \boldsymbol{\epsilon}_1,\end{aligned} \tag{22}$$

where $\boldsymbol{\epsilon}_1$ is assumed to be Gaussian white measurement noise. Further, we would like to parameterize the rotation matrix with respect to the Gibbs vector $\boldsymbol{a}_g = [g_1, g_2, g_3]^T$ to include the heading measurement in the error-state filter. Given the unit quaternion $\boldsymbol{q}_b^n = [q_1, q_2, q_3, q_4]^T$, the yaw angle can be obtained by (Fossen, 2021, p. 39)

$$\psi = \tan^{-1}\left(\frac{2(q_2q_3 + q_1q_4)}{1 - 2(q_3^2 + q_4^2)}\right). \tag{23}$$

We can express (23) with respect to the Gibbs vector by substituting the mapping $\boldsymbol{q}_b^n \to \boldsymbol{a}_g$, which follows from (11), such that

$$\begin{aligned}\psi &= \tan^{-1}\left(\frac{2q_1^2(g_1g_2 + g_3)}{1 - 2q_1^2(g_2^2 + g_3^2)}\right) \\ &= \tan^{-1}\left(\frac{2(g_1g_2 + g_3)}{1 + g_1^2 - g_2^2 - g_3^2}\right),\end{aligned} \tag{24}$$

where we have used the unit constraint, $(\boldsymbol{q}_b^n)^T(\boldsymbol{q}_b^n) = 1$, to obtain the last expression. Since we use the scaled Gibbs vector $\boldsymbol{a} = 2\boldsymbol{a}_g$, it follows that

$$\psi = h(\boldsymbol{a}) = \tan^{-1}\left(\frac{2(a_1a_2 + 2a_3)}{4 + a_1^2 - a_2^2 - a_3^2}\right), \tag{25}$$

where $\boldsymbol{a} = [a_1, a_2, a_3]^T$. By linearization about $\boldsymbol{a} = \hat{\boldsymbol{a}}$, we finally get

$$\delta y_\psi = \psi - h(\hat{\boldsymbol{a}}) \approx \frac{\partial h(\boldsymbol{a})}{\partial \boldsymbol{a}}^T\Big|_{a=\hat{a}}\delta\hat{\boldsymbol{a}}. \tag{26}$$

Eqs. (25) and (26) should be implemented using atan2, i.e., the four-quadrant inverse tangent confining the result to $[-\pi, \pi)$, to avoid singularities at $\psi = \pm\pi/2$. Note that the gradient used in the Kalman filter measurement matrix can also be computed from the unit quaternion:

$$\boldsymbol{c}_\psi(\hat{\boldsymbol{q}}_{\text{ins}}) := \frac{\delta h(\boldsymbol{a})}{\delta \boldsymbol{a}}\Big|_{\delta a=\delta\hat{a}}. \tag{27}$$

The error-measurement equations are summarized by (18), (19), (22) and (26). Following this order, the measurement vector used to correct the error estimates can be defined by

$$\delta \boldsymbol{y} = [(\delta \boldsymbol{y}_p)^T, (\delta \boldsymbol{y}_v)^T, (\delta \boldsymbol{v}_1)^T, \delta y_\psi]^T, \tag{28}$$

where $\delta \boldsymbol{y}_p$ denotes the measured position error, $\delta \boldsymbol{y}_v$ denotes the measured velocity error, $\delta \boldsymbol{v}_1$ denotes the error gravity reference vector, and δy_ψ is the measured heading error. For further details regarding the derivation of the error-measurement equations, we refer to Fossen (2021, p. 483–485).

2.5.3. Mathematical modeling of the error-state filter

Given the error-state equations (13)–(17) and the measurement vector (28), we proceed by presenting the mathematical models used for INS state propagation. The INS estimates are obtained by integrating the strapdown navigation equations with high-rate IMU measurements, i.e., specific force $\boldsymbol{f}_{nb}^b$ and angular velocity $\boldsymbol{\omega}_{nb}^b$. We emphasize that the strapdown navigation equations use the unit quaternion for attitude representation, while the MEKF uses the Gibbs vector. The INS state also includes the estimated sensor biases $\hat{\boldsymbol{b}}_{\text{ins,acc}}^b$ and $\hat{\boldsymbol{b}}_{\text{ins,gyro}}^b$ for online bias compensation since accelerometer and gyroscope biases will vary over time. This results in the following system of differential equations describing the INS estimates

$$\dot{\hat{\boldsymbol{p}}}_{\text{ins}}^n = \hat{\boldsymbol{v}}_{\text{ins}}^n \tag{29}$$

$$\dot{\hat{\boldsymbol{v}}}_{\text{ins}}^n = \boldsymbol{R}(\hat{\boldsymbol{q}}_{\text{ins}})\boldsymbol{f}_{\text{ins}}^b + \boldsymbol{g}^n \tag{30}$$

$$\dot{\hat{\boldsymbol{b}}}_{\text{ins,acc}}^b = \boldsymbol{0} \tag{31}$$

$$\dot{\hat{\boldsymbol{q}}}_{\text{ins}} = \boldsymbol{T}(\hat{\boldsymbol{q}}_{\text{ins}})\boldsymbol{\omega}_{\text{ins}}^b \tag{32}$$

$$\dot{\hat{\boldsymbol{b}}}_{\text{ins,gyro}}^b = \boldsymbol{0}, \tag{33}$$

where $\boldsymbol{f}_{\text{ins}}^b := \boldsymbol{f}_{\text{nb}}^b - \hat{\boldsymbol{b}}_{\text{ins,acc}}^b$ and $\boldsymbol{\omega}_{\text{ins}}^b := \boldsymbol{\omega}_{\text{nb}}^b - \hat{\boldsymbol{b}}_{\text{ins,gyro}}^b$ are the bias-compensated IMU measurements. Also, note that $\boldsymbol{T}(\hat{\boldsymbol{q}}_{\text{ins}})$ is a 4×3 quaternion transformation matrix from the body frame to the NED frame since $\hat{\boldsymbol{q}}_{\text{ins}} \equiv \hat{\boldsymbol{q}}_b^n$ and $\boldsymbol{g}^n = [0, 0, g(\mu)]^T$ is the WGS-84 ellipsoidal gravity vector.

One key point is that the INS estimates are reset by setting the estimated error state to zero for each incoming measurement, which is mathematically equivalent to $\hat{\boldsymbol{x}}_{\text{ins}} \leftarrow \hat{\boldsymbol{x}}_{\text{ins}} + \delta\hat{\boldsymbol{x}}$. Hence, the estimated error state $\delta\hat{\boldsymbol{x}}$ is computed by the MEKF before it is injected into the estimated INS state $\hat{\boldsymbol{x}}_{\text{ins}}$. Note that the unit quaternion $\hat{\boldsymbol{q}}_{\text{ins}}$ is rotated by the estimated unit quaternion error $\delta\hat{\boldsymbol{q}}_b^n$ using the Hamiltonian product, as shown in Algorithm 2. The error-state dynamics (13)–(17) can be represented by a 15-states model

$$\delta\dot{\boldsymbol{x}} = \boldsymbol{A}\delta\boldsymbol{x} + \boldsymbol{E}\boldsymbol{w} := \boldsymbol{f}(\delta\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}) \tag{34}$$

$$\delta\boldsymbol{y} = \boldsymbol{C}\delta\boldsymbol{x} + \boldsymbol{\epsilon} := \boldsymbol{h}(\delta\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{\epsilon}, \tag{35}$$

where $\boldsymbol{w} = [(\boldsymbol{w}_{\text{acc}}^b)^T, (\boldsymbol{w}_{b,\text{acc}}^b)^T, (\boldsymbol{w}_{\text{gyro}}^b)^T, (\boldsymbol{w}_{b,\text{gyro}}^b)^T]^T$ and $\boldsymbol{u} = [(\boldsymbol{f}_{nb}^b)^T, (\boldsymbol{\omega}_{nb}^b)^T, (\boldsymbol{g}^n)^T]^T$. After initialization, the nonlinear error-state model (34)–(35) is linearized about $\delta\boldsymbol{x}[k] = \boldsymbol{0}$ and $\delta\boldsymbol{w}[k] = \boldsymbol{0}$ for each time step k using Euler's integration method. Hence, we obtain the discretized error-state model

$$\delta\boldsymbol{x}[\text{k+1}] = \boldsymbol{A}_d[\text{k}]\delta\boldsymbol{x}[\text{k}] + \boldsymbol{E}_d[\text{k}]\boldsymbol{w}[\text{k}] \tag{36}$$

$$\delta\boldsymbol{y}[\text{k}] = \boldsymbol{C}_d[\text{k}]\delta\boldsymbol{x}[\text{k}] + \boldsymbol{\epsilon}[\text{k}] \tag{37}$$

with discrete-time system matrices

$$\begin{aligned}\boldsymbol{A}_d[k] &\approx \boldsymbol{I}_{15} + h\frac{\partial \boldsymbol{f}(\delta\boldsymbol{x}[k], \boldsymbol{u}[k], \boldsymbol{w}[k])}{\partial\delta\boldsymbol{x}[k]}\Big|_{\delta x[k]=0, \delta w[k]=0} \approx \boldsymbol{I}_{15} \\ &+ h\begin{bmatrix} \boldsymbol{0}_{3\times3} & \boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & -\boldsymbol{R}(\hat{\boldsymbol{q}}_{\text{ins}}[k]) & -\boldsymbol{R}(\hat{\boldsymbol{q}}_{\text{ins}}[k])\boldsymbol{S}(\boldsymbol{f}_{\text{ins}}^b[k]) & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & -\frac{1}{T_{\text{acc}}}\boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{S}(\boldsymbol{\omega}_{\text{ins}}^b[k]) & -\boldsymbol{I}_3 \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & -\frac{1}{T_{\text{gyro}}}\boldsymbol{I}_3 \end{bmatrix}\end{aligned} \tag{38}$$

$$\begin{aligned}\boldsymbol{C}_d[k] &\approx \frac{\partial \boldsymbol{h}(\delta\boldsymbol{x}[k], \boldsymbol{u}[k])}{\partial\delta\boldsymbol{x}[k]}\Big|_{\delta x[k]=0} \\ &\approx \begin{bmatrix} \boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{S}(\boldsymbol{R}^T(\hat{\boldsymbol{q}}_{\text{ins}}[k])\boldsymbol{v}_{01}^n) & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{1\times3} & \boldsymbol{0}_{1\times3} & \boldsymbol{0}_{1\times3} & \boldsymbol{c}_\psi^T(\hat{\boldsymbol{q}}_{\text{ins}}[k]) & \boldsymbol{0}_{1\times3} \end{bmatrix}\end{aligned} \tag{39}$$

$$\begin{aligned}\boldsymbol{E}_d[k] &\approx h\frac{\partial \boldsymbol{f}(\delta\boldsymbol{x}[k], \boldsymbol{u}[k])}{\partial\delta\boldsymbol{w}[k]}\Big|_{\delta x[k]=0, \delta w[k]=0} \\ &\approx h\begin{bmatrix} \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ -\boldsymbol{R}(\hat{\boldsymbol{q}}_{\text{ins}}[k]) & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & -\boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{I}_3 \end{bmatrix},\end{aligned} \tag{40}$$

where h is the IMU sampling time. Finally, we discretize the INS estimates (29)–(33) for the next time step k + 1. These are the estimates typically used by the vehicle for feedback control. We refer to Algorithm 2 for further details.

2.5.4. DVL-aided INS

The INS propagation is only driven by IMU-specific force and gyroscope measurements between aiding measurements from the camera. However, if a measurement dropout occurs, online IMU bias compensation is not possible due to the lack of absolute information from the aiding sensors. Unfortunately, the velocity and position integrated

Algorithm 2 Multiplicative extended Kalman filter

$\boldsymbol{Q}_d$, $\boldsymbol{R}_d$: Covariance matrices for the process and measurement noises
$\hat{\boldsymbol{P}}^-$, $\hat{\boldsymbol{P}}$: A priori and posterior covariance matrices
$\boldsymbol{K}$: Kalman gain
$\boldsymbol{g}^n$: WGS-84 ellipsoidal gravity vector as a function of latitude μ
$\boldsymbol{l}^b$: Lever-arm vector between the DVL and the body frame expressed in body frame
h: IMU sampling time

$\delta\hat{\boldsymbol{x}}[k] = [\delta\hat{\boldsymbol{p}}^n_{nb}[k]^T, \delta\hat{\boldsymbol{v}}^n_{nb}[k]^T, \delta\hat{\boldsymbol{b}}^b_{\text{acc}}[k]^T, \delta\hat{\boldsymbol{a}}[k]^T, \delta\hat{\boldsymbol{b}}^b_{\text{gyro}}[k]^T]^T$ (Error state)
$\hat{\boldsymbol{x}}_{\text{ins}}[k] = [\hat{\boldsymbol{p}}^n_{\text{ins}}[k]^T, \hat{\boldsymbol{v}}^n_{\text{ins}}[k]^T, \hat{\boldsymbol{b}}^b_{\text{acc,ins}}[k]^T, \hat{\boldsymbol{q}}_{\text{ins}}[k]^T, \hat{\boldsymbol{b}}^b_{\text{gyro,ins}}[k]^T]^T$ (INS state)

Input: $(h, \mu, \boldsymbol{f}^b_{\text{nb}}, \boldsymbol{\omega}^b_{\text{nb}}, \boldsymbol{y})$
Output: $\hat{\boldsymbol{x}}_{\text{ins}}$

$\hat{\boldsymbol{x}}_{\text{ins}}[0] = \boldsymbol{x}_0$, $\hat{\boldsymbol{P}}^-[0] = \boldsymbol{P}_0$, $\boldsymbol{Q}_d \leftarrow h\boldsymbol{Q}$, $\boldsymbol{R}_d \leftarrow h\boldsymbol{R}$, $\boldsymbol{g}^n \leftarrow [0, 0, g(\mu)]^T$ (Initialization)
for each new IMU message **do**
  $\boldsymbol{f}^b_{\text{ins}}[k] \leftarrow \boldsymbol{f}^b_{\text{nb}}[k] - \hat{\boldsymbol{b}}^b_{\text{acc,ins}}[k]$, $\boldsymbol{\omega}^b_{\text{ins}}[k] \leftarrow \boldsymbol{\omega}^b_{\text{nb}}[k] - \hat{\boldsymbol{b}}^b_{\text{gyro,ins}}[k]$ (Bias compensation)
  $\boldsymbol{A}_d \leftarrow \boldsymbol{I}_{15} + h\boldsymbol{A} + \frac{1}{2}(h\boldsymbol{A})^2$, $\boldsymbol{C}_d \leftarrow \boldsymbol{C}$, $\boldsymbol{E}_d \leftarrow h\boldsymbol{E}$ (Discrete-time matrices)
  if new measurement **then**
    if first measurement since dropout **then**
      $\hat{\boldsymbol{P}}^-[k] \leftarrow \boldsymbol{P}_0$ (Reset covariance matrix)
    end if
    $\boldsymbol{K}[k] \leftarrow \hat{\boldsymbol{P}}^-[k]\boldsymbol{C}_d^T[k](\boldsymbol{C}_d[k]\hat{\boldsymbol{P}}^-[k]\boldsymbol{C}_d^T[k] + \boldsymbol{R}_d[k])^{-1}$ (Kalman gain)
    $\delta\hat{\boldsymbol{x}}[k] \leftarrow \boldsymbol{K}[k](\boldsymbol{y}[k] - \boldsymbol{C}_d[k]\hat{\boldsymbol{x}}_{\text{ins}}[k])$ (Estimation error)
    $\Rightarrow \delta\hat{\boldsymbol{q}}^n_b[k] \leftarrow \frac{1}{\sqrt{4+\delta\hat{\boldsymbol{a}}[k]^T\delta\hat{\boldsymbol{a}}[k]}}\begin{bmatrix}2\\ \delta\hat{\boldsymbol{a}}[k]\end{bmatrix}$ (2 × Gibbs vector)
    $\hat{\boldsymbol{P}}[k] \leftarrow (\boldsymbol{I}_{15} - \boldsymbol{K}[k]\boldsymbol{C}_d[k])\hat{\boldsymbol{P}}^-[k](\boldsymbol{I}_{15} - \boldsymbol{K}[k]\boldsymbol{C}_d[k])^T + \boldsymbol{K}[k]\boldsymbol{R}_d[k]\boldsymbol{K}^T[k]$ (Corrector)

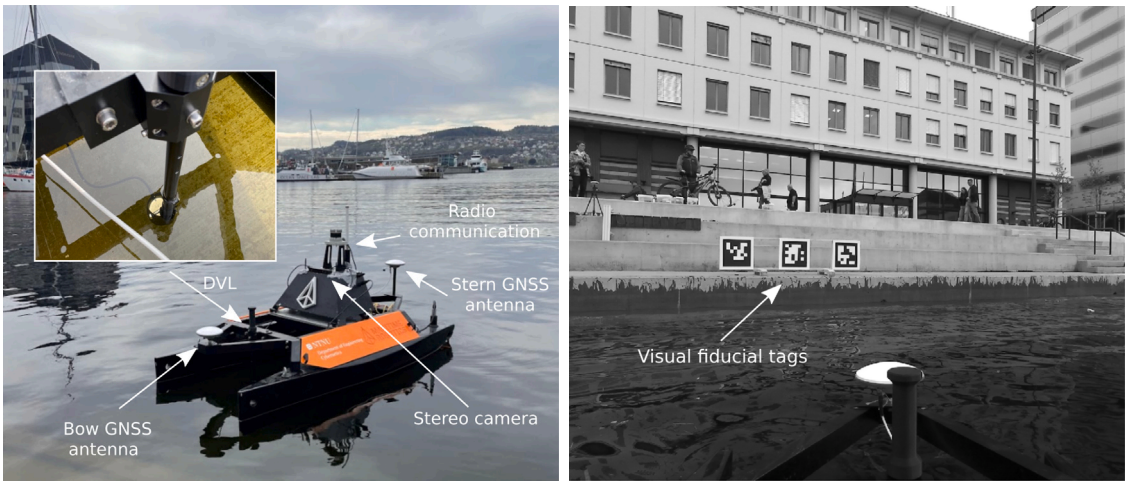
    // INS reset
    $\hat{\boldsymbol{x}}_{\text{ins}}[k] \leftarrow \hat{\boldsymbol{x}}_{\text{ins}}[k] + \delta\hat{\boldsymbol{x}}[k]$
    $\Rightarrow \hat{\boldsymbol{q}}_{\text{ins}}[k] \leftarrow \hat{\boldsymbol{q}}_{\text{ins}}[k] \otimes \delta\hat{\boldsymbol{q}}^n_{\text{nb}}[k]$ (Schur product)
    $\Rightarrow \hat{\boldsymbol{q}}_{\text{ins}}[k] \leftarrow \hat{\boldsymbol{q}}_{\text{ins}}[k]/||\hat{\boldsymbol{q}}_{ins}[k]||$ (Normalization)
  else
    $\hat{\boldsymbol{P}}[k] \leftarrow \hat{\boldsymbol{P}}^-[k]$ (No aiding)
  end if
  $\hat{\boldsymbol{P}}^-[k+1] \leftarrow \boldsymbol{A}_d[k]\hat{\boldsymbol{P}}[k]\boldsymbol{A}_d^T[k] + \boldsymbol{E}_d[k]\boldsymbol{Q}_d[k]\boldsymbol{E}_d^T[k]$ (Predictor)

  // INS propagation
  if measurement dropout **then**
    $\hat{\boldsymbol{p}}^n_{\text{ins}}[k+1] \leftarrow \hat{\boldsymbol{p}}^n_{\text{ins}}[k] + h\boldsymbol{R}^n_b\hat{\boldsymbol{q}}_{\text{ins}}[k](\boldsymbol{v}^{\text{dvl}}[k] + \boldsymbol{S}(\boldsymbol{\omega}^b_{\text{ins}}[k])\boldsymbol{l}^b)$ (DVL driven)
    $\hat{\boldsymbol{v}}^n_{\text{ins}}[k+1] \leftarrow \boldsymbol{R}^n_b\hat{\boldsymbol{q}}_{\text{ins}}[k](\boldsymbol{v}^{\text{dvl}}[k] + \boldsymbol{S}(\boldsymbol{\omega}^b_{\text{ins}}[k])\boldsymbol{l}^b)$ (DVL driven)
  else
    $\hat{\boldsymbol{p}}^n_{\text{ins}}[k+1] \leftarrow \hat{\boldsymbol{p}}^n_{\text{ins}}[k] + h\hat{\boldsymbol{v}}^n_{\text{ins}}[k]$ (Accelerometer driven)
    $\hat{\boldsymbol{v}}^n_{\text{ins}}[k+1] \leftarrow \hat{\boldsymbol{v}}^n_{\text{ins}}[k] + h(\boldsymbol{R}^n_b(\hat{\boldsymbol{q}}_{\text{ins}}[k])\boldsymbol{f}^b_{\text{ins}}[k] + \boldsymbol{g}^n)$ (Accelerometer driven)
  end if
  $\hat{\boldsymbol{q}}_{\text{ins}}[k+1] \leftarrow \hat{\boldsymbol{q}}_{\text{ins}}[k] + \exp^{T^n_b(h(\boldsymbol{\omega}^b_{\text{ins}}[k]))}\hat{\boldsymbol{q}}_{\text{ins}}[k]$ (Exact discretization)
  $\hat{\boldsymbol{q}}_{\text{ins}}[k+1] \leftarrow \hat{\boldsymbol{q}}_{\text{ins}}[k+1]/||\hat{\boldsymbol{q}}_{\text{ins}}[k+1]||$ (Normalization)
end for

from the accelerometer tend to drift very quickly without online accelerometer bias compensation, even with a fixed pre-compensated accelerometer bias. As such, we replace the accelerometer with low-drift velocity aiding from the DVL if a measurement dropout is detected. Hence, we integrate position directly from DVL velocity and assume constant DVL velocity between incoming DVL measurements, thus applying *nearest-neighbor interpolation*. This approach is preferred as long as the DVL outputs a relatively small amount of outliers. We emphasize that the AHRS inside the DVL enables enhanced velocity measurements, and outliers are rejected if not all of the beams have a line of sight to the sea bottom. Further, low drift induced by the gyroscope allows us to maintain a satisfactory attitude estimate without using the camera aiding for a period, i.e., a couple of minutes. This is very important not only for obtaining an accurate heading estimate but also because the DVL measurements must be transformed using the rotation matrix from body to NED, as shown in (5). Also, note that we reset the covariance matrix once we receive absolute corrections after a measurement dropout to avoid diverging INS estimates. This is because the open-loop estimates, without a reset, will lead to high uncertainty on the error state, thus blowing up the covariance matrix. For implementation details, we refer to Algorithm 2.

2.5.5. Filter tuning and initialization

The process and measurement noise covariance matrices, $\boldsymbol{Q}$ and $\boldsymbol{R}$, respectively, are the main tuning components for the Kalman filter. The diagonal entries of $\boldsymbol{Q}$, describing the uncertainty of the IMU's accelerometer and gyro and their associated biases, can be determined by the IMU random walk and in-run stability in the datasheet. Further, we discretize $\boldsymbol{Q}$ using a fixed IMU sampling time h. The measurement noise matrix $\boldsymbol{R}$, quantifying the uncertainty of the measurements, is



(a) (b)

Fig. 5. The experimental setup. (a) The Otter unmanned surface vehicle. (b) The tag configuration seen from the vehicles' camera view.

then tuned after the discretized process noise matrix $\boldsymbol{Q}_d$ is determined. We emphasize that $\boldsymbol{R}$ is tuned for a low ratio of DVL outliers. We use a synchronized dataset with experimental RTK INS data for validation. Hence, RTK INS is used to benchmark and evaluate the performance of the state estimates using root mean square error (RMSE). By evaluating different values of the diagonal entries of the discretized measurement noise matrix $\boldsymbol{R}_d$, we can obtain a low RMSE value. Finally, we test $\boldsymbol{Q}_d$ and $\boldsymbol{R}_d$ on different experimental data sequences to verify filter consistency. If landmark updates are not available, $\hat{\boldsymbol{x}}_{\text{ins}}$ is initialized using RTK INS estimates by converting longitude–latitude to North-East using (8) and (9). Since the estimated state is initialized using either accurate landmark observations or RTK INS, we assume that the estimated state is close to the true state initially. As a result, the initial covariance matrix $\hat{\boldsymbol{P}}^{-}[0]$ can be initialized with relatively low values on its diagonal entries.

3. Experimental setup

3.1. The USV platform

The Norwegian University of Science and Technology (NTNU) Otter is an USV produced by Maritime Robotics AS with two thrusters mounted at the stern. However, in-vehicle software and hardware were developed at the Department of Engineering Cybernetics, NTNU. The vehicle is equipped with a sensor suite that is used for experimental data acquisition in the harbor environment, as seen in Fig. 5(a). The experimental dataset is synchronized using the Precision Time Protocol, thus enabling accurate synchronization of sensor measurements across the network of devices onboard the USV.

3.1.1. Sensor configuration

The sensor suite consists of a Waterlinked A50 DVL (Waterlinked, 2022), a ZED 2i stereo camera (Stereolabs, 2022), an ADIS 16490 IMU (Analog Devices, 2022), and an SBG Ellipse 2D INS (Systems, 2022).

- **DVL:** The Waterlinked A50 DVL includes a 4-beam convex Janus array assisted by an in-built AHRS for enhanced accuracy. The sensor functions at an altitude level between 0.05 m and 50 m and at a maximum velocity of 3.75 m/s. Depending on the altitude level, the sensor outputs velocity estimates at a frequency between 4 and 15 Hz. The DVL is mounted on the vehicle such that each acoustic transducer has a line of sight to the sea bottom.
- **Camera:** We use the ZED 2i stereo camera with 2208×1242 pixel resolution and 110° horizontal FOV (per camera) and acquire the image data at a frequency of 15 Hz, the maximum for the selected resolution. Since color information is not necessary to detect the visual tags, we use the monochrome pixel format. As such, we also reduce bandwidth and processing while maintaining high image resolution for increased pose accuracy. Only the left camera is used since we employ monocular pose estimation.
- **IMU:** We use an ADIS 16490 IMU that is rigidly mounted to the center of the vehicle. It provides tri-axis angular rate measurements and tri-axis specific force measurements at a sample rate of 250 Hz. We refer to the datasheet (Analog Devices, 2022) for further details regarding the performance, including random walk and in-run stability of the sensor.
- **RTK INS:** The SBG Ellipse 2D INS is aided by raw GNSS data from stern and bow antennas and correction data from the RTK base at the dockside. After conducting an RTK survey for 16 hours, we achieved an absolute position precision of 5.6 cm. Using a dual-antenna configuration, the SBG Ellipse INS also obtains an accurate estimate of the heading of the USV. The INS estimates is expressed relative to the body frame for direct comparison to the proposed sensor fusion scheme.

3.2. Tag configuration

We use a tag configuration consisting of multiple tags for more accurate pose estimates, commonly referred to as a *tag bundle*. The tag bundle is used to extract a single pose from multiple tags rather than the poses of the individual tags. Hence, the pose estimation algorithm uses $4 \times n$ tag corners, where n is the number of detected tags. Our specific tag configuration employs three coplanar AprilTags of tag size 0.412 m $\times$ 0.412 m with fixed translations to each other. That is, they have a position offset of approximately 1 m to each other. The leftmost tag in Fig. 5(b) is the *reference tag*. Hence, the origin of the tag coordinate system is centered in the middle of the reference tag with axes defined according to the tag coordinate system in Fig. 2. Also, the tag IDs of interest are specified in a lookup table. As such, the visual fiducial system will only accept identifications that are defined in the lookup table, thus minimizing the number of false positives.

3.3. Experimental description

We conduct two experiments in a harbor environment under normal weather conditions with seabed conditions that are rocky and not

particularly muddy. The altitude level between the USV and the sea floor ranges from 0.2 m to 8 m throughout the experimental scenarios. Both experiments are performed offline on a high-performance laptop using experimental datasets from the harbor.

In Experiment 1, we assess the performance of open-loop estimates based on DVL and gyroscope measurements in three scenarios of the same duration, i.e., 3 minutes. In the first scenario, the DVL produces a relatively low rate of outliers throughout the sequence. In the second scenario, the DVL induces a relatively high rate of outliers in a period where the vehicle is moving at constant velocity. In the third scenario, the DVL produces a relatively high rate of outliers in a period where the vehicle is accelerating. The vehicle is operating at surge speeds between −0.5 m/s and 1 m/s, except for a short period in the third scenario where the vehicle exceeds 2 m/s in surge velocity. For all scenarios, the state estimates are compared to RTK INS measurements.

In Experiment 2, we extend the scope and investigate how the proposed scheme handles situations where the tags are too far from the vehicle to use landmark updates immediately, meaning that only acoustic and inertial measurements are available. Then, we investigate how the filter eliminates drift when the vehicle is inside the region where the tag measurements are accepted. In the last part of the sequence, we also explore how the method performs under two scenarios: When the vehicle has full access to visual measurements and when the vehicle is exposed to sensor dropout. The latter scenario is simulated by ignoring available tag measurements. It is, however, a realistic scenario since the USV is limited by a body-fixed camera FOV. The total sequence corresponds to the first 150 seconds of the third scenario in Experiment 1. For increased robustness, we still use DVL measurements in the measurement vector when the vehicle reaches the region where landmark measurements are accepted.

4. Experimental results

In Experiment 1, we present the results of the experiments by plotting the error in vehicle position and heading between the open-loop estimates and RTK INS. That is, we compute the 2-D Euclidean position error

$$e_{\mathrm{p}} = \sqrt{(x^{\mathrm{n}}_{\mathrm{rtk}} - \hat{x}^{\mathrm{n}}_{\mathrm{ins}})^2 + (y^{\mathrm{n}}_{\mathrm{rtk}} - \hat{y}^{\mathrm{n}}_{\mathrm{ins}})^2} \tag{41}$$

and the heading error

$$e_{\psi} = \psi^{n}_{\mathrm{rtk}} - \hat{\psi}^{n}_{\mathrm{ins}}, \tag{42}$$

where $\hat{\boldsymbol{p}}^{n}_{\mathrm{ins}} = [\hat{x}^{\mathrm{n}}_{\mathrm{ins}}, \hat{y}^{\mathrm{n}}_{\mathrm{ins}}, \hat{z}^{\mathrm{n}}_{\mathrm{ins}}]^T$ is the estimated position of the vehicle in the NED frame, $\hat{\psi}^{n}_{\mathrm{ins}}$ is the estimated heading angle relative to true north, $\boldsymbol{p}^{n}_{\mathrm{rtk}} = [x^{\mathrm{n}}_{\mathrm{rtk}}, y^{\mathrm{n}}_{\mathrm{rtk}}, z^{\mathrm{n}}_{\mathrm{rtk}}]^T$ is the RTK position of the vehicle in the NED frame, and ψ^{n}_{rtk} is the RTK heading angle relative to true north. In addition, associated surge and sway velocity estimated by the RTK INS and the DVL is presented. Note that only valid DVL measurements are shown. Finally, the accumulated number of DVL outliers compared to the total number of DVL measurements is shown.

In Experiment 2, we present the experimental results by plotting the USV position and heading, estimated by RTK INS, AprilTag, and the MEKF, when the USV slowly approaches the visual tags at the dockside before it reverses. The vehicle position is expressed in a local NED frame with origin in the reference tag, and the heading of the vehicle is expressed with respect to true north.

4.1. Experiment 1: DVL-aided INS

4.1.1. Scenario 1: Low rate of DVL outliers

The results from Scenario 1 in Experiment 1 are shown in Fig. 6. Fig. 6(d) shows that the DVL produces a relatively low amount of DVL outliers throughout the 3-minute sequence. The high number of valid DVL measurements, widely distributed over the sequence, are plotted together with ground truth surge and sway velocity in Fig. 6(c). As a result, the Euclidean position error e_p in Fig. 6(a) does not tend to accumulate, although the heading error e_ψ in Fig. 6(b) exceeds 2° several times during the first minute of the sequence.

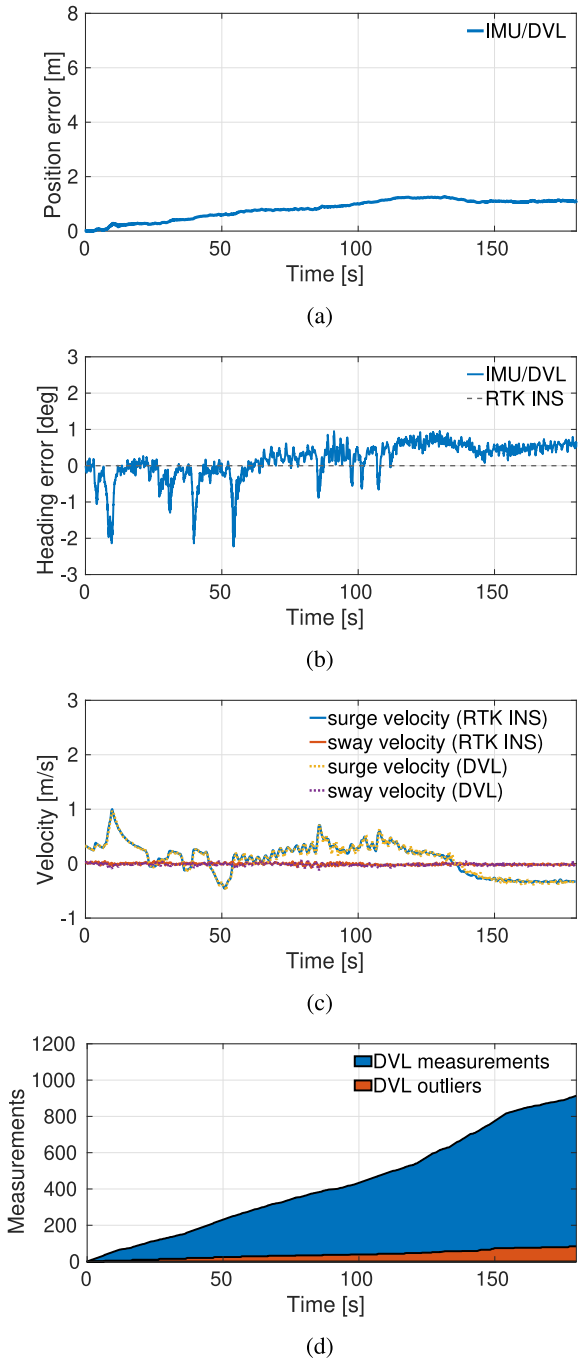


Fig. 6. The results from Scenario 1 in Experiment 1. (a) Position error between estimated position (IMU/DVL) and ground truth position (RTK INS). (b) Heading error between estimated position (IMU/DVL) and ground truth position (RTK INS). (c) Surge and sway velocity measured by the RTK INS and the DVL. (d) The accumulated number of DVL outliers compared to the total number of DVL measurements.

4.1.2. Scenario 2: DVL dropouts under constant velocity

The results from Scenario 2 in Experiment 1 are shown in Fig. 7. Fig. 7(d) shows that the DVL produces a relatively high amount of DVL outliers compared to the total amount of DVL measurements during the first 40 seconds of the sequence. In this period, the vehicle is mostly moving at constant surge and sway velocity, as seen in Fig. 7(c). As such, the resulting position error e_p accumulates more rapidly than in Scenario 1, as shown in Fig. 7(a).

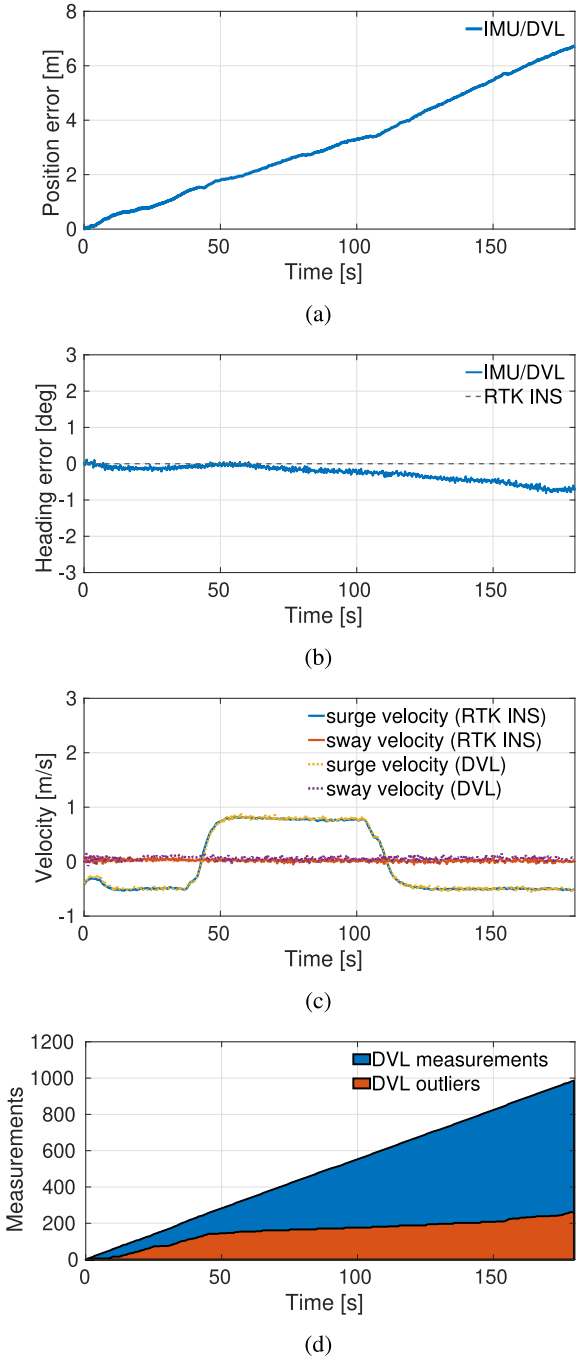


Fig. 7. The results from Scenario 2 in Experiment 1. (a) Position error between estimated position (IMU/DVL) and ground truth position (RTK INS). (b) Heading error between estimated position (IMU/DVL) and ground truth position (RTK INS). (c) Surge and sway velocity measured by RTK INS and DVL. (d) The accumulated number of DVL outliers compared to the total number of DVL measurements.

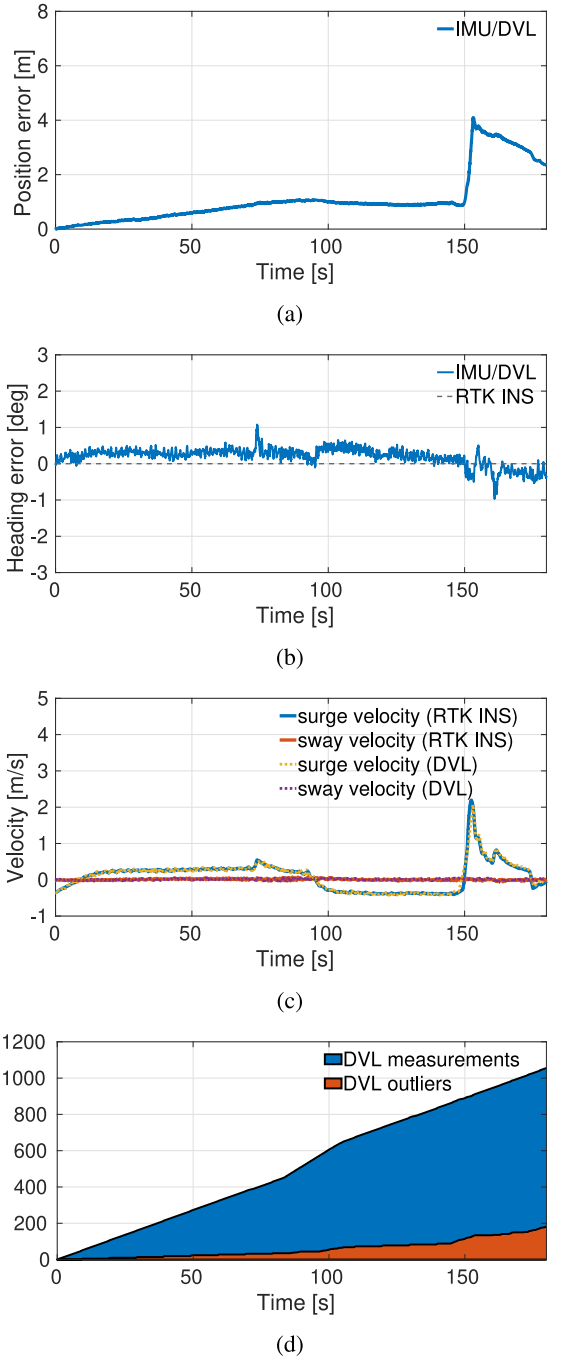


Fig. 8. The results from Scenario 3 in Experiment 1. (a) Position error between estimated position (IMU/DVL) and ground truth position (RTK INS). (b) Heading error between estimated position (IMU/DVL) and ground truth position (RTK INS). (c) Surge and sway velocity measured by the RTK INS and the DVL. (d) The accumulated number of DVL outliers compared to the total number of DVL measurements.

4.1.3. Scenario 3: DVL dropouts under acceleration

The results from Scenario 3 in Experiment 1 are shown in Fig. 8. Initially, the DVL produces a low rate of outliers resulting in a relatively low position error e_p. However, the DVL produces a large number of outliers in the last part of the sequence, as seen in Fig. 8(d). During this period, the vehicle accelerates significantly. As a result, the position error drifts very rapidly, as shown in Fig. 8(a). The fact that the position error tends to decrease after the sudden jump is likely a coincidence, as the high rate of DVL outliers during the last part of the sequence leads to high uncertainty regarding the position estimates.

4.2. Experiment 2: Camera-aided recovery

The results from Experiment 2 are shown in Figs. 9 and 10. The filter uses IMU/DVL as a navigation solution, thus producing INS estimates based on gyroscope and DVL measurements for the first 42 seconds of the sequence. We use Algorithm 1 with $d_{\text{thres}} = 20$ m to determine when the camera-tag pose measurements are accepted. Hence, the MEKF receives absolute corrections from the AprilTag system when the measured Euclidean distance between the reference tag and the camera is lower than 20 m and if at least two tags are detected

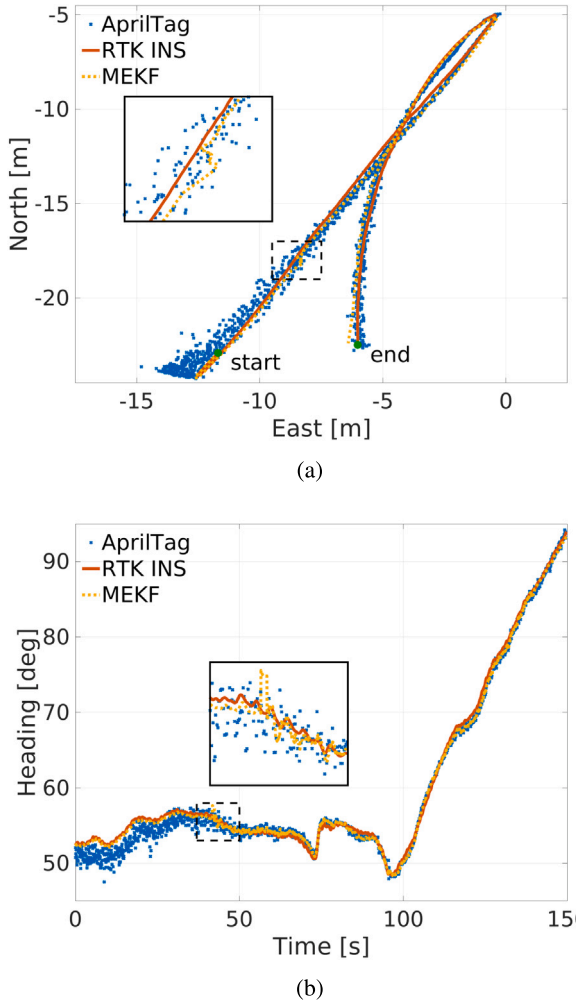


Fig. 9. The results from Experiment 2 with full access to tag measurements throughout the sequence. (a)–(b) Estimated position and heading of the vehicle using RTK INS, AprilTag, and the MEKF. The transition at which the MEKF starts to use position and heading measurements from the AprilTag system is shown in the zoomed plots.

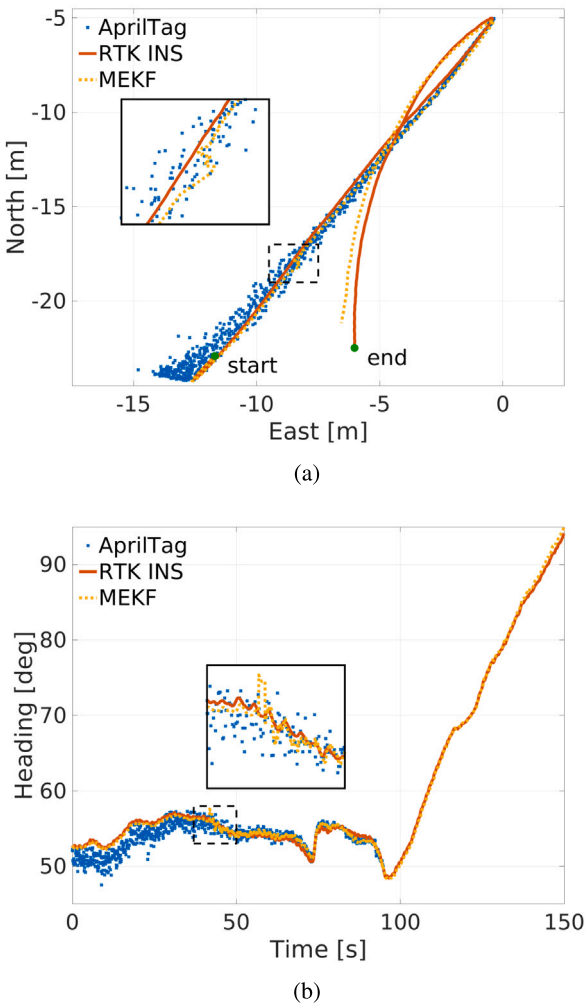


Fig. 10. The results from Experiment 2 with tag measurement dropout in the last part of the sequence. (a)–(b) Estimated position and heading of the vehicle using RTK INS, AprilTag, and the MEKF. The transition at which the MEKF starts to use position and heading measurements from the AprilTag system is shown in the zoomed plots.

simultaneously in an image. In the first scenario, the vehicle has full access to tag measurements throughout the sequence. The transition at which the MEKF starts to use position and heading measurements from the AprilTag system is shown in the zoomed plots in Figs. 9(a) and 9(b). As seen, the transition results in a small jump in position and heading followed by slightly more oscillating estimates compared to ground truth RTK INS.

In the second scenario, the vehicle is exposed to a sensor dropout in the last part of the sequence, resulting in slightly less accurate position estimates because the INS estimates are only based on gyroscope and DVL measurements under the dropout. Fortunately, the position error grows linearly and not quadratically since the DVL velocity measurement is integrated once to obtain the position. This can be seen in Fig. 10(a). The heading estimates, however, drift very slowly, as shown in Fig. 10(b).

4.3. Discussion of results

4.3.1. Experiment 1

In Experiment 1, we found that the performance of the IMU/DVL solution is heavily influenced by the rate of DVL outliers and the motion of the vehicle. For example, if the vehicle is exposed to high acceleration under a DVL dropout, the position error tends to accumulate much faster than if the vehicle moves with constant velocity. The vehicle is relatively small and, therefore, more exposed to pitching (up and down) than larger vehicles under acceleration. The varying pitching angle combined with a challenging seabed terrain, including rocks and partially sloping bottoms, may therefore result in a lost bottom fix for the DVL. This is likely to be the reason for the DVL dropout in the last part of Scenario 3, where the vehicle is accelerating. We also experienced that the gyroscope measurements did not accumulate large heading errors during the 3-minute sequences. As a result, the heading error did not influence the position error remarkably. This was also verified by replacing the DVL measurements with ground truth velocity measurements from the RTK INS, thus quantifying the position error induced by the gyroscope measurements. The outcome of the verification was expected due to the low velocity error between the DVL and RTK INS seen in Figs. 6(c), 7(c), and 8(c). These figures also confirm that the reduced performance in the presence of more DVL outliers is due to the reduced frequency of the inlying measurements and not the reduced quality of the inlying DVL measurements themselves.

Since the IMU/DVL solution results in very different positioning performances for sequences of the same length, it is important to discuss the metrics we can use to estimate how long the IMU/DVL solution provides satisfactory positioning performance. We argue that

acceleration should be taken into account when the DVL produces outliers. That is, we can penalize a DVL dropout under acceleration more than a DVL dropout under constant velocity since we assume constant velocity between the valid DVL measurements. This condition can, for example, be detected using IMU specific force measurements over a sliding window. Hence, if the vehicle is moving with constant velocity, we can sustain any number of outliers/dropouts without affecting the positioning performance remarkably. We also emphasize that the constant velocity model can be replaced with a constant acceleration model to improve position estimation under DVL dropouts. For long-endurance missions, the gyroscope measurements from the IMU will also generate position errors of significant magnitude, especially if the vehicle is exposed to high-speed maneuvers. In such situations, we recommend using an absolute heading reference, such as a gyrocompass, to eliminate the position error induced by the attitude sensor. As such, the accumulated position error is also less complicated to estimate since the positioning performance is solely dependent on the DVL.

4.3.2. Experiment 2

In Experiment 2, we extend the scope and demonstrate how the proposed MEKF can accurately estimate the full state of the vehicle in the harbor environment based on inertial measurements, low-drift acoustic measurements, and a drift-free camera solution aided by visual tags. In particular, we found that the proposed method was able to produce slowly drifting estimates based on DVL and gyroscope measurements and then eliminate the drift by using landmark updates from the AprilTags. This enables an enhanced operational range since the vehicle only needs to receive absolute corrections from the landmarks after some period, e.g., after a couple of minutes, as shown in Scenario 1 in Experiment 1. Since we assume ground truth measurements are not available, we cannot directly benchmark the performance of the visual fiducial system online. However, the experimental results show that the visual fiducial system produces sufficiently accurate measurements if we employ the criteria in Algorithm 1 with an appropriate d_{thres} value. Also, false positives are reduced to a minimum since the visual fiducial system only accepts detections with IDs pre-specified in the lookup table. As such, we avoid injecting outliers in the state estimator. Note that the accuracy of the visual fiducial system increases the closer the vehicle approaches the tags. As such, it is possible to adaptively change the measurement noise covariance matrix as a function of the distance to the landmarks. For simplicity, we have, however, avoided this modification.

Although the case study is limited to a harbor environment with a small set of visual tags, it is possible to extend the scope in which the vehicle can operate. For example, we can place landmarks in different regions to cover a larger area. As such, the vehicle can reset induced navigation errors when it is close to a set of landmarks and combine the DVL with an attitude sensor to produce low-drift estimates between each region of landmarks. Moreover, with a rotating camera or a multi-camera system for 360° view, the vision system is not limited by a body-fixed camera FOV and can, therefore, see landmarks in all directions. As such, the vehicle can receive landmark updates independently of the orientation of the vehicle, thereby increasing the flexibility of the path planner. This is particularly useful for underactuated vehicles. At last, we can add a camera with a larger optical zoom (at the cost of a narrower FOV) to detect landmarks at larger distances. Such a multi-camera system with different optical zoom and FOV can extend the range at which the vehicle can receive landmark updates.

5. Conclusions and future work

With recent advances in autonomy paving the way for increased use of autonomous and unmanned surface vehicles, employing independent and reliable alternatives to satellite-based navigation for redundancy is key to ensuring safe operation in harbor areas. To this end, we show how a strapdown INS aided by fiducial tags and acoustic measurements can be used to enhance the navigation capabilities of USVs in the harbor environment. Since the vehicle aid by fiducial tags that are range-sensitive and limited by the camera's FOV, we extend the operational scope at which the vehicle can operate by utilizing low-drift velocity aiding from the DVL. This was successfully demonstrated in Experiment 2, where the proposed sensor fusion scheme produced slowly drifting estimates based on DVL and gyroscope measurements before eliminating the drift using the visual fiducial system when the vehicle approached the dockside.

However, the proposed method is not trouble-free. For example, we found that the performance of the DVL-aided INS, without landmark updates, is heavily influenced by the motion of the vehicle and not necessarily the accumulated number of DVL outliers itself. In other words, if the vehicle is accelerating or progressing at approximately constant velocity when the DVL is exposed to outliers/dropouts is of great importance. For future work, we would, therefore, like to estimate an upper bound of time at which we trust the DVL-aided INS based on the variation of the DVL measurements near the outage or even vehicle actuation since actuation action is correlated to acceleration. As such, we can estimate the time (online) before a correction is needed, e.g., from landmarks or GNSS. We would also like to employ an absolute heading reference, such as a gyrocompass, to eliminate accumulating heading errors for long-endurance applications. Further, we plan to incorporate uncertainties from the landmark observations in the measurement model as a function of the camera-tag distance instead of assuming the same amount of Gaussian noise from the visual measurements that fulfills the proposed acceptance criteria. Finally, we would like to perform closed-loop field experiments, preferably with landmarks placed at different locations to cover larger areas in which the vehicle employs a GNSS-free navigation solution.

CRediT authorship contribution statement

Øystein Volden: Conceptualization, Methodology, Visualization, Software, Writing – original draft. **David Cabecinhas:** Supervision, Writing – review & editing. **António Pascoal:** Supervision, Writing – review & editing. **Thor I. Fossen:** Supervision, Resources, Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was funded by the Norwegian Research Council (project no. 223254) through the NTNU Center of Autonomous Marine Operations and Systems (AMOS) at the Norwegian University of Science and Technology. We are grateful to Waterlinked AS for providing the A50 DVL for field trials. We would also like to thank Petter Solnør for his assistance during the experimental data acquisition.

References

Analog Devices, 2022. ADIS16490. https://www.analog.com/media/en/technical-documentation/data-sheets/adis16490.pdf (Accessed: 08 Jun 2022).

Androjna, A., Brcko, T., Pavic, I., Greidanus, H., 2020. Assessing cyber challenges of maritime navigation. J. Mar. Sci. Eng. 8 (10), http://dx.doi.org/10.3390/jmse8100776, URL https://www.mdpi.com/2077-1312/8/10/776.

Aqel, M.O.A., Marhaban, M.H., Saripan, M.I., Ismail, N.B., 2016. Review of visual odometry: types, approaches, challenges, and applications. SpringerPlus 5, http://dx.doi.org/10.1186/s40064-016-3573-7.

Biswas, S.K., Qiao, L., Dempster, A.G., 2016. Computationally efficient unscented Kalman filtering techniques for launch vehicle navigation using a space-borne GPS receiver. In: Proceedings of the 29th International Technical Meeting of the Satellite Division of the Institute of Navigation. ION GNSS+ 2016, pp. 186–194.

Bolbot, V., Theotokatos, G., Boulougouris, E., Vassalos, D., 2020. A novel cyber-risk assessment method for ship systems. Saf. Sci. 131, 104908. http://dx.doi.org/10.1016/j.ssci.2020.104908, URL https://www.sciencedirect.com/science/article/pii/S0925753520303052.

Carroll, J.V., 2003. Vulnerability assessment of the U.S. transportation infrastructure that relies on the global positioning system. J. Navig. 56 (2), 185–193. http://dx.doi.org/10.1017/S0373463303002273.

Chen, J., Sun, C., Zhang, A., 2021. Autonomous navigation for adaptive unmanned underwater vehicles using fiducial markers. In: 2021 IEEE International Conference on Robotics and Automation. ICRA, pp. 9298–9304. http://dx.doi.org/10.1109/ICRA48506.2021.9561419.

Collins, T., Bartoli, A., 2014. Infinitesimal plane-based pose estimation. Int. J. Comput. Vis. 109 (3), 252–286.

Crassidis, J.L., Markley, F.L., Cheng, Y., 2007. Survey of nonlinear attitude estimation methods. J. Guid. Control Dyn. 30 (1).

DNV, 2018. Autonomous and remotely operated ships. URL https://rules.dnv.com/docs/pdf/DNV/cg/2018-09/dnvgl-cg-0264.pdf (Accessed: 03 June 2022).

Farell, J.A., Barth, M., 1998. The Global Positioning System and Inertial Navigation. McGraw-Hill.

Farrell, J., 2008. Aided Navigation: GPS with High Rate Sensors. McGraw-Hill, New York, NY.

Fiala, M., 2005. ARTag, a fiducial marker system using digital techniques. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2. CVPR'05, pp. 590–596.

Fossen, T.I., 2021. Handbook of Marine Craft Hydrodynamics and Motion Control, 2nd Ed. Wiley.

Fossen, T.I., Perez, T., 2004. Marine Systems Simulator (MSS). https://github.com/cybergalactic/MSS.

Fukuda, G., Hatta, D., Guo, X., Kubo, N., 2021. Performance evaluation of IMU and DVL integration in marine navigation. Sensors 21 (4), http://dx.doi.org/10.3390/s21041056, URL https://www.mdpi.com/1424-8220/21/4/1056.

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., Marín-Jiménez, M., 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognit. 47 (6), 2280–2292. http://dx.doi.org/10.1016/j.patcog.2014.01.005, URL https://www.sciencedirect.com/science/article/pii/S0031320314000235.

Hsu, H.Y., Toda, Y., Yamashita, K., Watanabe, K., Sasano, M., Okamoto, A., Inaba, S., Minami, M., 2022. Stereo-vision-based AUV navigation system for resetting the inertial navigation system error. Artif. Life Robot. 1–14.

Hussain, A., Akhtar, F., Khand, Z.H., Rajput, A., Shaukat, Z., 2021. Complexity and limitations of GNSS signal reception in highly obstructed enviroments. Eng. Technol. Amp Appl. Sci. Res. 11 (2), 6864–6868. http://dx.doi.org/10.48084/etasr.3908, URL https://www.etasr.com/index.php/ETASR/article/view/3908.

Julier, S., Uhlmann, J., 2004. Unscented filtering and nonlinear estimation. Proc. IEEE 92 (3), 401–422. http://dx.doi.org/10.1109/JPROC.2003.823141.

Kalaitzakis, M., Cain, B., Carroll, S., Ambrosi, A., Whitehead, C., Vitzilaios, N., 2021. Fiducial markers for pose estimation. J. Intell. Robot. Syst. 101 (4), 1–26.

Kato, H., Billinghurst, M., 1999. Marker tracking and HMD calibration for a video-based augmentedreality conferencing system. In: The 2nd International Workshop on Augmented Reality. IWAR 99, pp. 85–94. http://dx.doi.org/10.1109/IWAR.1999.803809.

Kretschmann, L., Burmeister, H.-C., Jahn, C., 2017. Analyzing the economic benefit of unmanned autonomous ships: An exploratory cost-comparison between an autonomous and a conventional bulk carrier. Res. Transp. Bus. Manag. 25, 76–86. http://dx.doi.org/10.1016/j.rtbm.2017.06.002, New developments in the Global Transport of Commodity Products URL https://www.sciencedirect.com/science/article/pii/S2210539516301328.

Malyuta, D., Brommer, C., Hentzen, D., Stastny, T., Siegwart, R., Brockers, R., 2020. Long-duration fully autonomous operation of rotorcraft unmanned aerial systems for remote-sensing data acquisition. J. Field Robotics 37 (1), 137–157. http://dx.doi.org/10.1002/rob.21898, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/rob.21898.

Malyuta, D., Wolfgang, M., 2022. Apriltag_ros. https://github.com/AprilRobotics/apriltag_ros.

Markley, F.L., Crassidis, J.L., 2014. Fundamentals of Spacecraft Attitude Determination and Control, Vol. 1286. Springer.

Mondjar-Guerra, V., Garrido-Jurado, S., Muoz-Salinas, R., Marn-Jimnez, M.J., Medina-Carnicer, R., 2018. Robust identification of fiducial markers in challenging conditions. Expert Syst. Appl. 93 (C), 336–345. http://dx.doi.org/10.1016/j.eswa.2017.10.032.

Mularie, W., 2000. World geodetic system 1984–its definition and relationships with local geodetic systems. Department of Defense, NIMA USA.

Myint, M., Yonemori, K., Lwin, K.N., Yanou, A., Minami, M., 2018. Dual-eyes vision-based docking system for autonomous underwater vehicle: An approach and experiments. J. Intell. Robot. Syst. 92 (1), 159–186. http://dx.doi.org/10.1007/s10846-017-0703-6.

Olson, E., 2011. AprilTag: A robust and flexible visual fiducial system. In: 2011 IEEE International Conference on Robotics and Automation. pp. 3400–3407.

Prochniewicz, D., Grzymala, M., 2021. Analysis of the impact of multipath on Galileo system measurements. Remote Sens. 13 (12), http://dx.doi.org/10.3390/rs13122295, URL https://www.mdpi.com/2072-4292/13/12/2295.

Relling, T., Lützhöft, M., Ostnes, R., Hildre, H.P., 2018. A human perspective on maritime autonomy. In: Schmorrow, D.D., Fidopiastis, C.M. (Eds.), Augmented Cognition: Users and Contexts. Springer International Publishing, Cham, pp. 350–362.

Romanovas, M., Ziebold, R., Lança, L., 2015. A method for IMU/GNSS/Doppler velocity log integration in marine applications. In: 2015 International Association of Institutes of Navigation World Congress. IAIN, pp. 1–8. http://dx.doi.org/10.1109/IAIN.2015.7352241.

ROS, 2022. ROS - Robot Operating System. https://ros.org/.

Rubiola, E., 2008. Phase Noise and Frequency Stability in Oscillators. Cambridge University Press.

Rudolph, D., Wilson, T.A., 2012. Doppler velocity log theory and preliminary considerations for design and construction. In: 2012 Proceedings of IEEE Southeastcon. pp. 1–7. http://dx.doi.org/10.1109/SECon.2012.6196918.

Salavasidis, G., Munafò, A., Harris, C.A., Prampart, T., Templeton, R., Smart, M., Roper, D.T., Pebody, M., McPhail, S.D., Rogers, E., Phillips, A.B., 2019. Terrain-aided navigation for long-endurance and deep-rated autonomous underwater vehicles. J. Field Robotics 36 (2), 447–474. http://dx.doi.org/10.1002/rob.21832, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/rob.21832.

Sola, J., 2017. Quaternion kinematics for the error-state Kalman filter. arXiv preprint arXiv:1711.02508.

Stereolabs, 2022. ZED 2i. https://www.stereolabs.com/assets/datasheets/zed-2i-datasheet-feb2022.pdf (Accessed: 13 July 2022).

Systems, S.B.G., 2022. Ellipse series. https://www.sbg-systems.com/products/ellipse-series/ellipse-d_rtk_gnss_ins (Accessed: 04 June 2022).

Trslic, P., Rossi, M., Robinson, L., O'Donnel, C.W., Weir, A., Coleman, J., Riordan, J., Omerdic, E., Dooly, G., Toal, D., 2020. Vision based autonomous docking for work class ROVs. Ocean Eng. 196, 106840. http://dx.doi.org/10.1016/j.oceaneng.2019.106840, URL https://www.sciencedirect.com/science/article/pii/S0029801819309369.

Vargas, E., Scona, R., Willners, J.S., Luczynski, T., Cao, Y., Wang, S., Petillot, Y.R., 2021. Robust underwater visual SLAM fusing acoustic sensing. In: 2021 IEEE International Conference on Robotics and Automation. ICRA, pp. 2140–2146. http://dx.doi.org/10.1109/ICRA48506.2021.9561537.

Volden, Ø., 2022. Visual-acoustic state estimation. https://github.com/oysteinvolden/dvl_state_estimation.

Volden, Ø., Stahl, A., Fossen, T.I., 2021. Vision-based positioning system for auto-docking of unmanned surface vehicles (USVs). Int. J. Intell. Robot. Appl. 6 (1), 86–103.

Wang, J., Olson, E., 2016. AprilTag 2: Efficient and robust fiducial detection. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, pp. 4193–4198.

Waterlinked, 2022. The world's smallest DVL – DVL A50. https://www.waterlinked.com/hubfs/Product_Assets/DVL_A50/wl-21035-3_DVL_A50.pdf?hsLang=en (Accessed: 15 Sept 2022).

Xu, S., Luczynski, T., Willners, J.S., Hong, Z., Zhang, K., Petillot, Y.R., Wang, S., 2021. Underwater visual acoustic SLAM with extrinsic calibration. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, pp. 7647–7652. http://dx.doi.org/10.1109/IROS51168.2021.9636258.

Ziebold, R., Medina, D., Romanovas, M., Lass, C., Gewies, S., 2018. Performance characterization of GNSS/IMU/DVL integration under real maritime jamming conditions. Sensors 18 (9), http://dx.doi.org/10.3390/s18092954, URL https://www.mdpi.com/1424-8220/18/9/2954.