

RESEARCH ARTICLE

An Enhanced Navigation System With Predictive Motion Planning for Unmanned Surface Vehicles in GNSS-Attenuated Dynamic Urban Waterways

Jiarong Liu^{1,2}  | Mingyang Li^{1,2} | Hong Liu³ | Lijian Wan³ | Jinbo Chen⁴ | Yongsheng Zhao^{1,2}

¹School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China | ²State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai, China | ³Shanghai Dahua Surveying & Mapping Technology Co. Ltd., Shanghai, China | ⁴School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China

Correspondence: Yongsheng Zhao (yongsheng@sjtu.edu.cn)

Received: 25 September 2024 | **Revised:** 14 April 2025 | **Accepted:** 31 May 2025

Funding: This study was funded by National Natural Science Foundation of China (No. 52271284) and the Oceanic Interdisciplinary Program of Shanghai Jiao Tong University (No. SL2021ZD201).

Keywords: autonomous navigation | dynamic obstacle avoidance | multisensor fusion | simultaneous localization and mapping | unmanned surface vehicles

ABSTRACT

Unmanned surface vehicles (USVs) applied in urban waterways may suffer from inaccurate localization due to the Global Navigation Satellite System (GNSS) attenuation, and be susceptible to collision threats from vessels of human-induced violations and piloting errors. This paper proposes an enhanced navigation framework capable of stable continuous localization, dynamic obstacle perception, and collision-free motion planning. A tightly coupled LiDAR-Visual-Inertial Odometry via Smoothing and Mapping (LVI-SAM) is selected as the fundamental framework of localization and mapping subsystem. An incrementally mapping data structure is incorporated to improve the computation efficiency and accuracy of the LiDAR odometry optimization process. To mitigate the long-term accumulating odometry drift, valid GNSS measurements are introduced to provide absolute reference in the factor graph optimization framework, which can achieve optimum state estimation by maximum a posteriori given all the noisy measurements from multiple sensors. Furthermore, a dynamic occupancy grid map framework, based on sequential Monte Carlo and probability hypothesis density method, is developed to enhance situational awareness of USVs for risk anticipation of dynamic obstacles and facilitate predictive avoidance. Extensive real-world experiments have been carried out to demonstrate that the proposed autonomous navigation system is capable of robust and accurate localization over long-term urban waterway navigation, and dynamic obstacle avoidance through a safer predictive strategy.

1 | Introduction

Unmanned surface vehicles (USVs) have gained widespread adoption due to their operational flexibility and payload capacity across various maritime domains, ranging from offshore exploration (Fun Sang Cepeda et al. 2023) to environmental monitoring (Ramkumar Sudha et al. 2024), and maritime emergency response operations. In recent years, the growing demand for urban waterway applications, including urban transportation

(Wang et al. 2023), surface waste collection (Li et al. 2023), bathymetry for dredging planning (Zhao et al. 2024), pollutant surveillance, and so on, has spurred increased research interest in autonomous USV technologies tailored for inland water scenarios. USVs equipped with a sophisticated autonomous navigation system showcase a promising role in sustainable economy development of cities with ample waterway networks. However, urban water environments present distinct challenges compared with open-sea conditions. USVs operating in these confined

waterways frequently encounter Global Navigation Satellite System (GNSS) degradation or loss, compounded by the complex navigation requirements of dynamic obstacle avoidance in narrow, vessel-dense environments, posing great challenges to USV autonomous navigation.

Autonomous localization with high precision and high reliability stands as a pivotal technique within the autonomous navigation system for USVs, especially in the congested and unstructured urban waterways. The localization methods deployed in marine environments typically integrate inertial measurement units (IMUs) with GNSS receivers using Kalman-filter-based fusion strategies (Cole and Schamberg 2022; W. Liu et al. 2019; W. Liu, Liu, et al. 2023) and high-precision positioning methods based on real-time kinematics (RTK) (Xiang et al. 2023), may encounter adaptability challenges and thus invalid estimates when applied in urban waterway environments. The reason lies in the potential obstruction of satellite signals by buildings and bridges (Cheng et al. 2021), as well as the occurrence of multipath effects, where the receiver detects indirect signals reflected by the ground or other terrain obstacles.

To address diverse GNSS-degraded scenarios, researchers have proposed real-time localization techniques based on various sensors within the framework of simultaneous localization and mapping (SLAM), aiming to enable a range of autonomous downstream tasks. For large-scale environments such as harbors, Zhou et al. (2021) developed a real-time three-dimensional (3D) light detection and ranging (LiDAR) SLAM system that utilized coarse-to-fine scan matching, and Normal Distribution Transform process accelerated by a dynamic voxel grid storage structure, achieving global drift correction at a level below 1%. To enable berthing operations of USVs in restricted waters, a customized berth assistance system was developed by H. Wang, Yin, et al. (2024). This system integrated SLAM algorithms fusing LiDAR and RTK data. In the context of narrow inland water scenarios, Y. Wang, Liu, et al. (2024) utilized high-planarity points for pose estimation in LiDAR-based SLAM and introduced a rejection sampling method to eliminate points with negligible contribution to pose estimation updates, thus improving the accuracy of generalized iterative closest point (ICP) matching. Additionally, several studies have explored the availability of visual SLAM on USVs. For example, Volden et al. (2022) designed a low-cost stereo-vision-based positioning system to complement GNSS–IMU navigation, providing relative positioning estimates during the terminal phase of autonomous docking. The visual fiducial system, utilizing artificial tag detection, was further enhanced by integrating a Doppler velocity log (DVL) sensor and an IMU within the multiplicative expanded Kalman filter (Volden et al. 2023). X. Liu, Hu, et al. (2023) proposed a visual-inertial odometry system capable of extracting point and line features from bridge ceilings as landmarks, enabling stable estimation of USV state variables in GPS-denied environments.

Notably, single-sensor degradation may occur due to the unstable motion of USVs caused by irregular wind-wave disturbances, unstructured landmark characteristics, water reflections, and other challenging conditions in adverse urban waterways. While the aforementioned studies have been validated in numerous confined scenarios, they may face single-sensor degradation

issues in more general urban waterway environments, particularly when USVs are deployed for long-term operational tasks in complex scenes. Therefore, a multimodal sensor fusion scheme, which is driven by LiDAR, visual, IMU, GNSS, and other types of measurements, manifests a more viable and promising solution.

However, research on factor-graph-based LiDAR-visual-inertial (LVI) SLAM systems for USV long-term operation in complex urban waterways remains limited. Thoms et al. (2023) enhanced an LVI system by incorporating a preintegrated velocity factor derived from DVL acoustic measurements and applied it to a USV designed for infrastructure inspection. However, their work only included a real-world experiment in a single limited scenario under a water-spanning bridge. In Wang et al. (2019), LiDAR, IMU, and camera data were integrated within an extended Kalman filter (EKF) framework for USV state estimation in urban waterways, with LiDAR serving as the primary sensor. Nevertheless, detailed experiments and evaluations to assess the individual contribution of each sensor to localization performance were not conducted. This prompts the first motivation for this study, which focuses on the development of a robust multisensor SLAM system for USV long-term operation in urban waterways.

Dynamic obstacle avoidance in narrow urban waterways presents another distinct technical challenge for the autonomous and intelligent operation of USVs. This challenge stems from two key factors. First, urban waterways are typically cluttered with unpredictable dynamic obstacles (e.g., drifting debris and vessels) and static boundary constraints (e.g., embankments and bridges), significantly challenging LiDAR-based perception systems. These challenges are exacerbated by the inherent sparsity of LiDAR point clouds in large-scale aquatic environments and persistent noise artifacts caused by wave reflection. Second, the maneuverability constraints of USVs (D. Liu et al. 2022), combined with the difficulty of incorporating predictive information about dynamic obstacles (Zhu and Ding 2023), pose significant challenges for local reactive avoidance algorithms. Recent progress in addressing collision avoidance problems in such complex scenarios remains limited. To enable USV autonomous navigation in complex urban canals, Shan, Wang, et al. (2020) proposed a receding horizon reactive planner that hierarchically evaluated collision risk, turning cost, and distance cost during path replanning. Gonzalez-Garcia et al. (2022) addressed the path-following and obstacle-avoidance problem within a nonlinear model predictive control framework, where cluttered obstacles detected through LiDAR measurements were mathematically modeled as a series of nonlinear circular inequality constraints. To achieve robust collision avoidance for USVs during path-following tasks, a high-order control barrier function-based approach leveraging the elliptical shape representation for the purpose of reducing the safety conservatism was proposed in Wen et al. (2024). Li and Zhang (2022) enhanced the velocity obstacle algorithm for USVs to ensure safe navigation in the emergency case that dynamic obstacle ships may temporarily turn while adhering to the International Regulations for Preventing Collisions at Sea (COLREGS). Similarly, Han et al. (2022) proposed a guidance strategy combining COLREGS-compliant repulsive potential field with a stationary grid map, in adaptation to a complex scene of multi-USVs.

Most existing methods perform reactive collision avoidance using only instantaneous sensor measurements or assume predefined velocity information of dynamic obstacles in simulated environments. Additionally, representing static and dynamic obstacles using regular geometric shapes (e.g., circles and ellipses) typically necessitates preprocessing steps of raw sensor data, like, clustering and matching. However, this strategy struggles to handle obstacles with arbitrary geometries or large-scale structures. Most importantly, due to the USV's inherent inertia and controller latency, dynamic obstacle avoidance cannot be resolved by merely increasing path replanning rates. Consequently, predicting dynamic obstacles with filtered velocity information becomes essential. These limitations motivate this study to develop a local reactive planner that operates on a map representation capable of precisely encoding complex static and dynamic obstacles with their velocities.

Distinct from previous work, this study advances the field of autonomous navigation for USVs in GNSS-degraded urban waterways by addressing two real-world critical challenges: long-term consistent localization and safety-guaranteed navigation amidst dynamic obstacles. The key innovations are summarized as follows:

1. A heterogeneous multisensor fusion framework for long-term low-drift USV navigation in urban waterways is proposed, integrating LiDAR, monocular camera, IMU, and sparse GNSS measurements within a factor graph optimization architecture to improve continuity and consistency of localization for USVs in GNSS-degraded scenarios. An incrementally updated map data structure is introduced to eliminate the need for map rebuilding before the scan-to-map matching process in the LiDAR odometry submodule, enabling higher efficiency. GNSS measurements are strategically fused to globally bound drift via the smoothness and convergence properties of factor graph optimization.
2. A novel predictive kinematic-compliant reactive planner that is capable of handling collision avoidance confronted with dynamic obstacles in a predictive manner is proposed. On the basis of the sequential Monte Carlo and probability hypothesis density (SMC-PHD) filter, a dynamic occupancy grid map for decision making is developed, enabling filtered velocity estimation and probabilistic occupancy predictions of dynamic obstacles in arbitrary shapes. Motion constraints explicitly incorporating minimum turning radius and acceleration bounds of USVs' kinematic are embedded during the replanning process.
3. Extensive field tests demonstrate the capability of our localization system: sustaining submeter accuracy under intermittent GNSS outages and loop closure after long-term navigation. Obstacle-avoidance reliability is also validated by long-term navigation and dynamic avoidance functionality by the complex experimental scenarios through interactions with dynamic obstacles. The whole system can operate without prior maps, making it deployable in unmapped or rapidly changing environments.

The rest of this paper is organized as follows. Section 2 presents the problem formulation of localization and collision avoidance

of USV in urban waterways. Section 3 details the methodology of multisensor-based localization and mapping, particle-based dynamic occupancy prediction, and dynamic-constraint compliant path planning. Hardware setup and experiments are presented in Section 4. Section 5 concludes the paper.

Notation. \mathbb{R}^n denotes the n -dimensional Euclidean space. $SO(3) \in \mathbb{R}^{3 \times 3}$ denotes the special orthogonal group, which represents rotation transformations in 3D space. $(3) = SO(3) \times \mathbb{R}^3$ denotes the special Euclidean group, representing rigid body transformations that include both rotation and translation. A vector is denoted by a bold lowercase letter (e.g., \mathbf{v}), while a matrix with a bold uppercase letter (e.g., \mathbf{T}). Let $\tilde{\mathbf{A}}$ denote the noisy measurement of A . A coordinate frame is represented by a hand-written uppercase letter, such as \mathcal{W} , which can also be viewed as a group belonging to \mathbb{R}^n . ${}^W\mathbf{v}$ represents a vector defined in the coordinate frame. ${}^L\mathbf{T} \in SE(3)$ denotes the transformation that maps a vector from frame L to frame \mathcal{W} .

2 | Problem Formulation

This study addresses two critical challenges in USV navigation within urban waterways: continuous localization and autonomous collision avoidance. Both tasks are intrinsically linked to a unified coordinate frame known as the east-north-up (ENU) frame. The ENU frame is defined such that its X-axis points east, its Y-axis points north, and its Z-axis points to the zenith. Once the initial latitude, longitude, and altitude are established as the reference origin, the local ENU frame can be established. This local ENU frame serves as the default world frame, denoted as \mathcal{W} .

2.1 | Localization

The system state estimated in the localization task is defined as

$$\mathbf{x} \triangleq [{}^W\mathbf{R}, {}^W\mathbf{p}, {}^W\mathbf{v}, \mathbf{b}_a, \mathbf{b}_\omega], \quad (1)$$

where ${}^W\mathbf{R} \in SO(3)$ is the orientation of the USV body frame \mathcal{B} in the frame \mathcal{W} , ${}^W\mathbf{p} \in \mathbb{R}^3$ is the translational vector, and ${}^W\mathbf{v} \in \mathbb{R}^3$ is the linear velocity of \mathcal{B} in \mathcal{W} . The IMU acceleration and angular velocity biases are expressed as $\mathbf{b}_a \in \mathbb{R}^3$ and $\mathbf{b}_\omega \in \mathbb{R}^3$.

Traditionally, an EKF framework to fuse IMU and GNSS measurements is employed to estimate the system state. This framework operates in a two-step cycle:

1. *Prior estimation:* The state is predicted through discretized integration of IMU measurements.
2. *Posterior correction:* The predicted state is updated using valid GNSS positioning data.

However, in areas with GNSS outages, this approach accumulates significant drift errors due to the reliance on pure integration of noisy IMU measurements. Over time, these errors render the state estimation unreliable.

To address this issue, the system states are estimated from continuous multimodal measurements from heterogeneous sensors, including a LiDAR, a monocular camera, an IMU, and a GNSS receiver. LiDAR provides a set of 3D measurement points $\mathcal{P} = \{\mathbf{p} \in \mathbb{R}^3\}$ in its local coordinate \mathcal{L} by scanning the surroundings at a fixed frequency. Monocular camera captures two-dimensional (2D) RGB images, which cannot be directly used as landmarks unless a 2D point set $\mathcal{F} = \{\mathbf{f} \in \mathbb{R}^2\}$ in its local frame \mathcal{C} is extracted through visual feature detection algorithms. IMU delivers high-frequency, discretized measurements of linear acceleration $\tilde{\mathbf{a}} \in \mathbb{R}^3$, angular velocity $\tilde{\boldsymbol{\omega}} \in \mathbb{R}^3$ in its local coordinate \mathcal{B} and orientation ${}^W_B\mathbf{R}$ in the local ENU frame. The IMU frame is rigidly aligned with the USV body frame to establish a unified reference for motion sensing and control. GNSS receiver provides global positioning data in the earth-fixed frame, including latitude, longitude, and altitude, which are transformed via the local ENU conversion to offer absolute position references that are crucial for correcting localization drift errors. Under the assumption of a fixed rigid connection between the sensors, all multimodal measurements can be transformed into the body frame \mathcal{B} , jointly facilitating accurate state estimation of the USV.

Figure 1 illustrates a scenario in which a USV navigates under a bridge arch. In such a case, GNSS signals are blocked, but the LiDAR provides geometric scanning information of the structure, while the camera extracts feature points from the ceiling surface. When the USV moves from position \mathcal{B}_{k-1} to \mathcal{B}_k , two consecutive sets of LiDAR point clouds (\mathcal{P}_{k-1} and \mathcal{P}_k), or camera feature points (\mathcal{F}_{k-1} and \mathcal{F}_k) are generated and then utilized to construct an optimization problem. This problem minimizes the sum of point-to-plane or reprojection residuals by adjusting the $SE(3)$ transformation between two adjacent sets. These methods can be interpreted as LiDAR and visual odometry, which inherently exhibit far lower drift error compared with IMU odometry. Therefore, our objective is to

leverage LiDAR and monocular camera data to detect landmarks from structures that cause GNSS outages, thereby maintaining reliable USV state estimation.

2.2 | Collision Avoidance

Local reactive avoidance algorithm typically makes decisions based on newly acquired perception data of dynamic obstacles. This approach works efficiently if the vehicle is sufficiently agile. Figure 2 illustrates the trajectory replanning process when encountering a dynamic obstacle without prediction. At time $t = k - 1$, a trajectory is replanned, but by time $t = k$, it becomes occupied, necessitating the generation of a new feasible trajectory. This switching process can be tedious and challenging for the USV controller. This underscores the need to incorporate dynamic obstacle prediction into the collision avoidance process.

Let the configuration space within which the USV operates be defined as \mathcal{Q} , and let $\mathcal{Q}_{\text{obs}} \subset \mathcal{Q}$ denote the subset space occupied by stationary and dynamic obstacles. A trajectory is parameterized by a normalized scalar $\tau \in [0,1]$ and defined as $\sigma(\tau) \subset \mathcal{Q}$. The collision avoidance problem, which aims to find a feasible and optimal path from the USV current position ${}^W\mathbf{p}_{\text{cur}}$ to the goal ${}^W\mathbf{p}_{\text{goal}}$, can be formulated as

$$\sigma^* = \underset{\sigma \in \mathcal{Q}_{\text{free}}}{\operatorname{argmin}} [c_s(\sigma) + c_d(\sigma) + c_r(\sigma)] \quad (2a)$$

$$\text{s.t. } \sigma(0) = {}^W\mathbf{p}_{\text{cur}}, \quad (2b)$$

$$\sigma(1) = {}^W\mathbf{p}_{\text{goal}}, \quad (2c)$$

where $\mathcal{Q}_{\text{free}}$ represents the traversable space within \mathcal{Q} , excluding the space \mathcal{Q}_{obs} occupied by obstacles. The cost functions $c_s(\sigma)$,

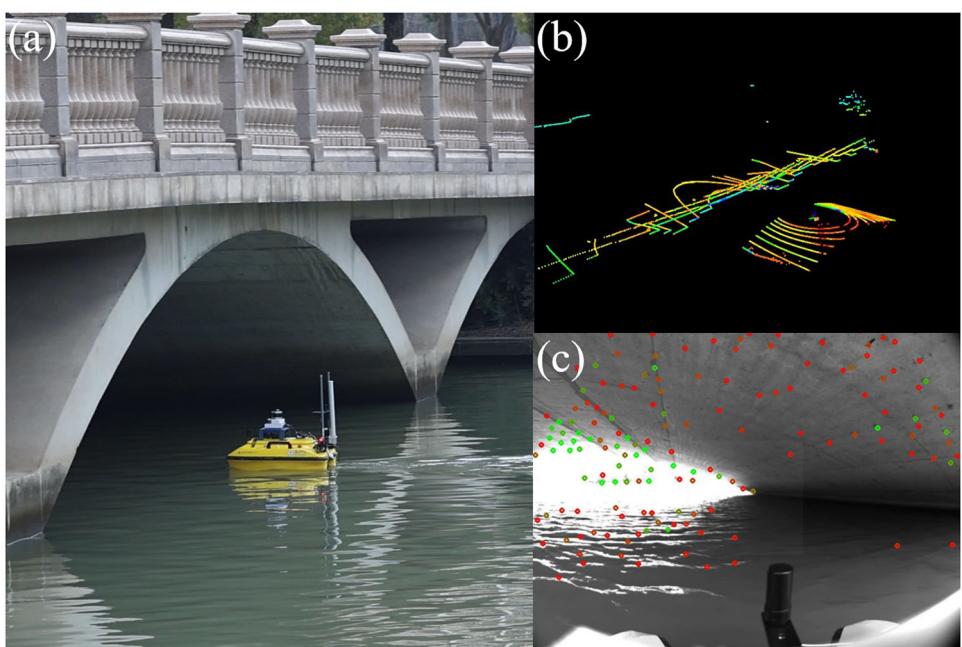


FIGURE 1 | A typical GNSS-attenuation scene of an urban waterway. (a) A USV navigates through a bridge arch, (b) a LiDAR scan of the bridge structure, and (c) an image with feature extraction of the arch ceiling. GNSS, Global Navigation Satellite System; LiDAR, light detection and ranging; USV, unmanned surface vehicle. [Color figure can be viewed at wileyonlinelibrary.com]

$c_d(\sigma)$, and $c_r(\sigma)$ evaluate the smoothness, sailing distance, and collision risk of a trajectory, respectively. To enable predictive navigation, \mathcal{Q}_{obs} is further divided into $\mathcal{Q}_{\text{obs},\text{cur}}$ and $\mathcal{Q}_{\text{obs},\text{pred}}$, representing the current and anticipated obstacle distribution in the future, respectively. To model the distribution of static and dynamic obstacles, we employ a 2D dynamic occupancy grid map. Additionally, we utilize an SMC-PHD filter to estimate the occupancy probabilities and velocity states of each grid. This estimation is based on consecutive LiDAR point sets and their corresponding poses derived from the localization framework described earlier.

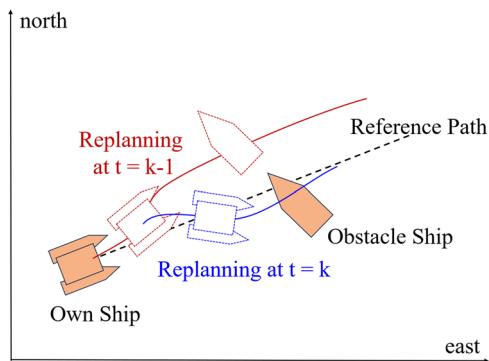


FIGURE 2 | A trajectory replanning process encountering a dynamic obstacle without prediction. [Color figure can be viewed at wileyonlinelibrary.com]

3 | Methodology

3.1 | System Pipeline of USV Navigation

The architecture of the proposed framework for autonomous USV navigation in dynamic urban waterways is illustrated in Figure 3. The developed framework consists of two core components: multisensor fusion localization module and kinematic-compliant predictive collision avoidance module. Initially, the multisensor fusion localization module achieves continuous and accurate USV state estimation by integrating data from multiple sources, including LiDAR point clouds, visual features extracted from camera images, motion measurements from IMU, and global positioning data from a GNSS receiver, detailed in Section 3.2. Then, the successive sets of LiDAR point clouds, along with their corresponding pose estimates, are further fed into the SMC-PHD filter to detect and predict the distribution of surrounding obstacles in the form of a dynamic occupancy grid map, which is introduced in Section 3.3.1 in detail. This obstacle awareness is critical for the subsequent decision-making process of collision avoidance. Finally, Section 3.3.2 introduces the collision avoidance module operating in a two-step scheme: the front-end kinematic-compliant, grid-based path searching to identify initial feasible paths, and the back-end trajectory optimization using B-spline control points to improve the smoothness, safety distance, and dynamic feasibility. To construct the distance cost function, a 2D Euclidean Signed Distance Field (ESDF) is generated based on the

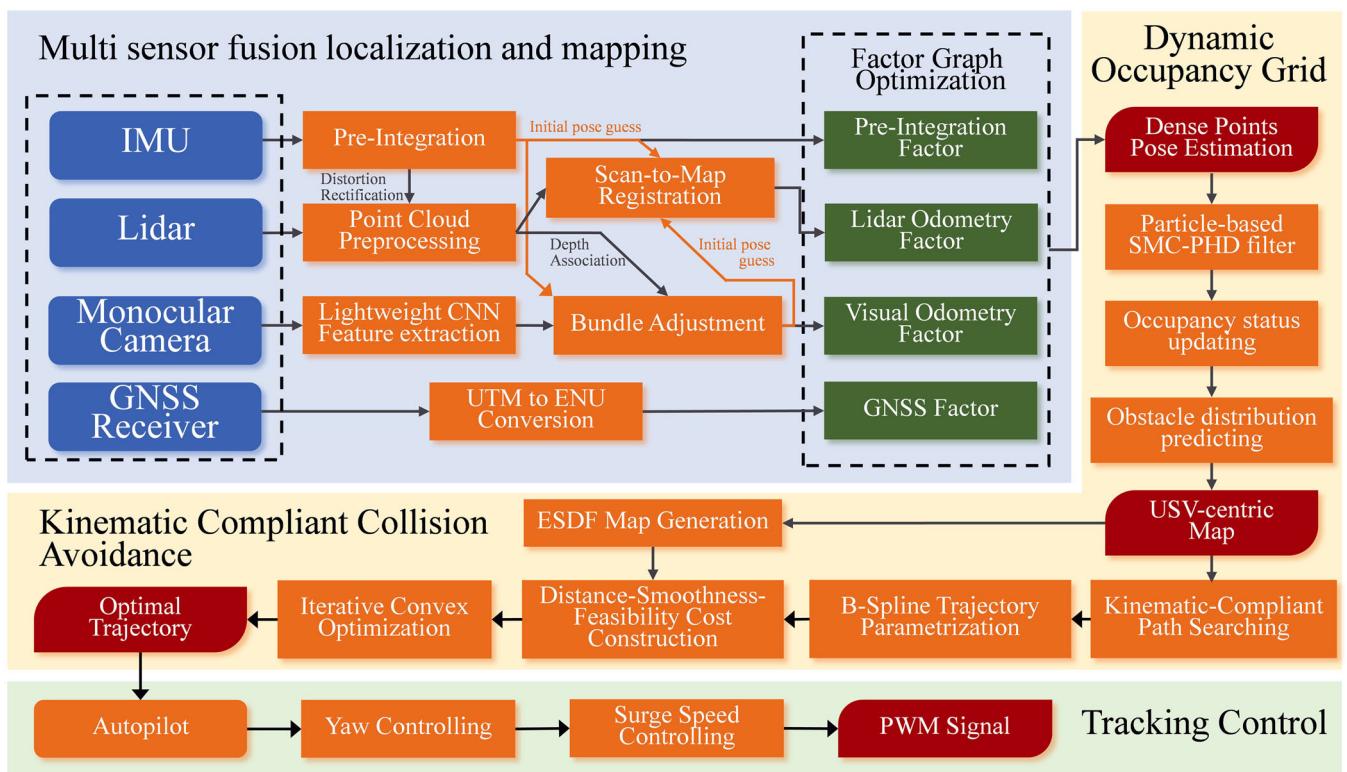


FIGURE 3 | Enhanced navigation system framework for USVs in urban waterways. It contains three layers: localization and mapping, collision avoidance, and tracking control. CNN, convolutional neural network; ENU, east-north-up; ESDF, Euclidean Signed Distance Field; GNSS, Global Navigation Satellite System; IMU, inertial measurement unit; LiDAR, light detection and ranging; PHD, probability hypothesis density; PWM, pulse-width modulation; SMC, sequential Monte Carlo; USV, unmanned surface vehicle; UTM, Universal Transverse Mercator. [Color figure can be viewed at wileyonlinelibrary.com]

predicted dynamic occupancy grid map. Through bilinear interpolation, the ESDF provides both the distances for cost calculation and the corresponding distance gradients, which are essential for optimizing the trajectory.

3.2 | Multisensor Fusion Localization

In this study, we formalize and develop the multisensor fusion localization module tailored for aquatic environments, building upon the state-of-the-art framework LiDAR-Visual-Inertial Odometry via Smoothing and Mapping (LVI-SAM) (Shan et al. 2021). LVI-SAM integrates two complementary odometry systems: LiDAR-inertial system (LIS) from LIO-SAM (Shan, Englot, et al. 2020), and visual-inertial system (VIS) from VINS-MONO (T. Qin et al. 2018). This combination enables the framework to handle challenging and more sensor degradation scenarios effectively. Both subsystems operate in real-time and can function independently, ensuring robustness even if one system fails. LVI-SAM is built with a factor graph optimization architecture (Dellaert and Kaess 2017), which inherently supports the integration of diverse sensor measurements. This capability enhances both the accuracy and robustness of the localization system. Given these advantages, we adopt LVI-SAM as a solid baseline to achieve smooth and globally consistent localization in complex urban waterways.

3.2.1 | LiDAR-Inertial System

At the data preprocess stage, LIS begins by leveraging high-frequency IMU measurements to correct distortions in LiDAR point clouds caused by the USV's unstable six-degree-of-freedom motion on the disturbed water surface. To streamline this process, we adopt the IMU preintegration technique proposed in Christian et al. (2015), which avoids computationally expensive repeated integrations over all IMU measurements. This method calculates the relative motion transformations between timestamps t_i and t_j , offering two key advantages: serving as constraints between adjacent states in the factor graph, and rectifying motion-induced distortions in LiDAR point clouds. The relative increments of position, velocity, and orientation are inferred using Euler integration as follows:

$$\overset{\mathcal{B}_i}{\mathbf{B}_j} \mathbf{R} = \prod_{k=i}^{j-1} \text{Exp}((\tilde{\omega}_k - \mathbf{b}_\omega - \mathbf{n}_\omega) \Delta t), \quad (3a)$$

$$\overset{\mathcal{B}_i}{\mathbf{B}_j} \mathbf{v} = \sum_{k=i}^{j-1} [\overset{\mathcal{B}_i}{\mathbf{B}_k} \mathbf{R} (\tilde{\mathbf{a}}_k - \mathbf{b}_a - \mathbf{n}_a) \Delta t], \quad (3b)$$

$$\overset{\mathcal{B}_i}{\mathbf{B}_j} \mathbf{p} = \sum_{k=i}^{j-1} \left[\frac{3}{2} \cdot \overset{\mathcal{B}_i}{\mathbf{B}_k} \mathbf{R} (\tilde{\mathbf{a}}_k - \mathbf{b}_a - \mathbf{n}_a) \Delta t^2 \right]. \quad (3c)$$

Here, k represents the index of IMU measurements, and Δt is the constant sampling interval between consecutive measurements. During the initial preintegration phase, the IMU biases \mathbf{b}_a , \mathbf{b}_ω , as well as the discrete-time noise terms \mathbf{n}_a , \mathbf{n}_ω are temporarily ignored. These factors are only accounted for after the factor graph optimization finishes. $\text{Exp}(*)$ is the exponential map defined as follows:

$$\text{Exp}: \phi \in \mathbb{R}^3 \rightarrow \exp(\phi^\wedge) \in SO(3), \quad (4a)$$

$$\phi^\wedge = \begin{bmatrix} 0 & -\phi_3 & \phi_2 \\ \phi_3 & 0 & \phi_1 \\ -\phi_2 & -\phi_1 & 0 \end{bmatrix}. \quad (4b)$$

To simplify the explanation of the distortion rectification process, consider a LiDAR point ${}^{\mathcal{L}} \mathbf{p}$ sampled at the same timestamp as the k th IMU measurement. This point can be mapped to the reference frame \mathcal{B}_i (corresponding to the start of the scan) using the following transformation:

$${}^{\mathcal{L}_i} \mathbf{p}_{\text{undistort}} = {}^{\mathcal{L}_i} \mathbf{R} \left({}^{\mathcal{B}_k} \mathbf{R} \cdot {}^{\mathcal{L}} \mathbf{R} \cdot {}^{\mathcal{L}_i} \mathbf{p} + {}^{\mathcal{B}_k} \mathbf{p} \right), \quad (5)$$

where ${}^{\mathcal{L}} \mathbf{R}$ represents the extrinsic parameter matrix that reprojects LiDAR points from the LiDAR frame \mathcal{L} to the IMU frame \mathcal{B} , and ${}^{\mathcal{L}} \mathbf{R}$ is its inverse. By applying this transformation to each point in the LiDAR scan, all points are reprojected in a unified frame, thus effectively completing the rectification.

In the subsequent feature extraction step, the undistorted LiDAR points are classified into edge features and planar features based on a local smoothness metric (Zhang and Singh 2017). However, in urban waterways, which often involve large-scale environments, the sparsity of LiDAR points at long distances from reference structures makes it challenging to extract stable edge features. These edge features are often too noisy to elevate the accuracy of the scan-to-map registration process. As highlighted in Z. Liu, Liu, et al. (2023), an environment with sufficient edge features typically also provides an adequate number of planar features for robust scan-to-map registration. Therefore, we retain the feature extraction step but focus on filtering out noisy edge features and retaining stable planar features for registration.

After extracting planar features from the i th LiDAR scan, which are denoted as ${}^{\mathcal{L}_i} \widetilde{\mathcal{P}}_{\text{planar}}$, we first transform them from \mathcal{L}_i to \mathcal{W} according to the initial pose guess from IMU preintegration or VIS. Then we obtain the point set ${}^{\mathcal{W}} \widetilde{\mathcal{P}}_{\text{planar},i}$ and match it with the global map point set \mathcal{M}_i . Different from the original scheme, \mathcal{M}_i are now constructed by incrementally adding the historical keyframe LiDAR scans up to time $i-1$. This implementation is achieved through the use of an incremental K-dimensional-Tree (iKd-Tree), a binary search tree designed to dynamically insert and delete 3D points (Cai et al. 2021). The iKd-Tree significantly enhances the smoothness and accuracy of the LIS, which will be demonstrated in Section 4.2.

To achieve a balance between map density and memory consumption, the keyframe selection strategy is exploited. A new keyframe is added only when the vehicle's movement exceeds a user-defined threshold. This strategy ensures that the most relevant LiDAR points and their associated poses are retained, effectively reducing redundancy while maintaining high precision in localization and mapping. Upon this, we obtain an incrementally updating \mathcal{M}_i to perform efficient accurate nearest neighbor searches (Friedman et al. 1977) for each point in the current LiDAR scan ${}^{\mathcal{W}} \widetilde{\mathcal{P}}_{\text{planar},i}$. This process, known as correspondence matching, identifies a set of points $\mathbf{p}_{\mathcal{M},(l,m,n)} \in \mathcal{M}_i$ that are closest to each

query point $\tilde{\mathbf{p}}_u \in {}^W\widetilde{\mathcal{P}}_{\text{planar},i}$, where u, l, m, n are the feature indices in their corresponding sets. On the basis of this correspondence, a point-to-plane residual is formulated as follows:

$$d_{\text{planar},u} = \frac{|(\tilde{\mathbf{p}}_u - \mathbf{p}_{M,l}) \cdot ((\mathbf{p}_{M,l} - \mathbf{p}_{M,m}) \times (\mathbf{p}_{M,l} - \mathbf{p}_{M,n}))|}{|(\mathbf{p}_{M,l} - \mathbf{p}_{M,m}) \times (\mathbf{p}_{M,l} - \mathbf{p}_{M,n})|}. \quad (6)$$

Then an optimization problem for obtaining the optimal LiDAR pose in local map frame can be defined by summing up the point-to-plane residuals:

$$\min_{{}^W\widetilde{\mathbf{T}}_i} \left(\sum_{\tilde{\mathbf{p}}_u \in {}^W\widetilde{\mathcal{P}}_{\text{planar},i}} d_{\text{planar},u}(\tilde{\mathbf{p}}_u, {}^W\widetilde{\mathbf{T}}_i) \right), \quad (7)$$

where ${}^W\widetilde{\mathbf{T}}_i \triangleq \begin{bmatrix} {}^W\widetilde{\mathbf{R}}_i & {}^W\widetilde{\mathbf{p}}_i \end{bmatrix} \in (3)$ is the estimated LiDAR pose at the beginning of the i th scan in the local map frame \mathcal{W} . However, in challenging scenarios characterized by geometric self-symmetry or featureless environments, such as long canals, it becomes difficult to aggregate sufficient point-to-plane constraints from LiDAR scans. This can lead to an ill-conditioned optimization problem, causing degeneracy in certain directions of the LiDAR pose. This issue manifests as the LiDAR-slip phenomenon, where the estimated pose drifts in underconstrained directions.

To address this, we adopt a state-of-the-art method proposed in Tuna et al. (2022) to distinguish between localizable and non-localizable directions. This localizability-aware framework can operate in parallel with the point-to-plane correspondence procedure. By performing singular value decomposition (SVD) and eigen-analysis on the translation and rotation Hessian matrices of Equation (7) we obtain the eigenvectors representing the directions of localization. Subsequently, a pointwise analysis of localizability contribution is conducted for each registered point in the scan. This analysis quantifies the localizability of each direction, corresponding to the aforementioned eigenvectors. Due to space limitations, detailed derivations are not included here. Ultimately, the scan-to-map registration can be reformulated as a constrained optimization problem, using the eigenvectors to constraining the updating of the LiDAR pose in nonlocalizable directions. To address the problem in a tractable formulation, the LiDAR pose should be parameterized in 6D using the logarithmic mapping, then the Equation (7) is rederived as

$$\min_{\tilde{\xi}_i} \left(\sum_{\tilde{\mathbf{p}}_u \in {}^W\widetilde{\mathcal{P}}_{\text{planar},i}} d_{\text{planar},u}(\tilde{\mathbf{p}}_u, \tilde{\xi}_i) \right) \quad (8a)$$

$$\text{s.t. } \mathbf{C}\tilde{\xi}_i - \mathbf{d} = 0, \quad (8b)$$

$$\tilde{\xi}_i = \log({}^W\widetilde{\mathbf{T}}_i) = \begin{bmatrix} \boldsymbol{\phi}_i \\ \boldsymbol{\rho}_i \end{bmatrix} \in \mathbb{R}^6, \quad (8c)$$

where $\boldsymbol{\phi}_i$ and $\boldsymbol{\rho}_i$ are the rotational and translational vectors, respectively, \mathbf{C} is the matrix composed of the transpose of

eigenvectors corresponding to nonlocalizable directions in a row-major manner, \mathbf{d} is the projection vector of first iterated $\tilde{\xi}_{i,\text{iter}=0}$ onto these nonlocalizable directions. \mathbf{C} and \mathbf{d} are calculated as follows:

$$\mathbf{C} = \begin{bmatrix} \boldsymbol{\eta}_{\text{nloc},j} & \mathbf{0}_{m_\phi \times 3} \\ \vdots & \vdots \\ \mathbf{0}_{m_\rho \times 3} & \boldsymbol{\eta}_{\text{nloc},j} \end{bmatrix} \in \mathbb{R}^{(m_\phi + m_\rho) \times 6}, \quad (9a)$$

$$\mathbf{d} = \begin{bmatrix} \boldsymbol{\eta}_{\text{nloc},j} \cdot \boldsymbol{\phi}_{i,\text{iter}=0} \\ \vdots \\ \boldsymbol{\eta}_{\text{nloc},j} \cdot \boldsymbol{\rho}_{i,\text{iter}=0} \end{bmatrix} \in \mathbb{R}^{(m_\phi + m_\rho) \times 1}, \quad (9b)$$

where $\boldsymbol{\eta}_{\text{nloc},j}$ is the transpose of the eigenvectors corresponding to nonlocalizable directions (abbreviated as nloc), m_ϕ and m_ρ denote the number of rotational and translational constraints, respectively. This constrained optimization approach can be solved via SVD after augmentation of linear equal constraints with Lagrangian multipliers. This offers the advantage of mitigating the adverse effects of nonlocalizable directions on the localizable ones during scan-to-map registration iteration, and maintains the initial pose estimation from VIS in the non-localizable directions.

3.2.2 | Visual Inertia System

VIS tightly coupling the monocular visual measurements and IMU measurements in the bundle adjustment formulation. It maximizes the likelihood of the observed measurements to reach maximum a posteriori (MAP) estimation of the history states within a sliding window. The corresponding augmented state vector of the sliding window is defined as

$$\begin{aligned} \boldsymbol{\chi} &= [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n, \boldsymbol{\rho}_0, \boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_l, \dots, \boldsymbol{\rho}_m] \mathbf{x}_i \\ &= \left[{}^W\mathbf{r}_i, {}^W\mathbf{p}_i, {}^W\mathbf{v}_i, \mathbf{b}_a, \mathbf{b}_\omega \right], \quad i \in [0, n], l \in [0, m], \end{aligned} \quad (10)$$

here n is the number of keyframes in the sliding window, and m is the number of features, and $\boldsymbol{\rho}_l$ is the inverse depth of the feature l from its first observation, which is initialized by triangulation of structure from motion (SFM). The depth estimation of the feature can be better initialized through feature depth association with a local map composed of the several recent dense LiDAR scans (Zhang and Singh 2018). By aggregating the prior residual \mathbf{r}_{pr} , visual reprojection residuals \mathbf{r}_c , and preintegrated IMU measurements residuals \mathbf{r}_i , the visual-inertial odometry problem is formulated as the following nonlinear least-square problem:

$$\begin{aligned} \boldsymbol{\chi}^* &= \arg \max_{\boldsymbol{\chi}} p(\boldsymbol{\chi}) \prod p(Z|\boldsymbol{\chi}) = \arg \min_{\boldsymbol{\chi}} \\ &\left\{ \| \mathbf{r}_{\text{pr}} - \mathbf{H}_{\text{pr}} \boldsymbol{\chi} \|^2 + \sum_{i=0}^{n-1} \left\| \mathbf{r}_i \left(\widetilde{\mathbf{Z}}_{C_{i+1}}^{C_i}, \boldsymbol{\chi} \right) \right\|_{\Sigma_{C_{i+1}}^{C_i}}^2 \right. \\ &\left. + \sum_{j \in [0, n]} \left\| \mathbf{r}_c \left(\widetilde{\mathbf{Z}}_1^{C_j}, \boldsymbol{\chi} \right) \right\|_{\Sigma_c}^2 \right\}, \end{aligned} \quad (11)$$

where $\|\cdot\|_{\Sigma}$ is the Mahalanobis norm with covariance Σ , and $\{\mathbf{r}_{\text{pr}}, \mathbf{H}_{\text{pr}}\}$ encapsulates the prior information of the states via the Schur complement marginalization. $\tilde{\mathcal{Z}}_{\mathcal{C}_{i+1}}^{\mathcal{C}_i}$ is the preintegration of IMU measurements between image \mathcal{C}_i and \mathcal{C}_{i+1} , and $\tilde{\mathcal{Z}}_l^j$ represents the measurement of feature l in frame j . Detailed calculation of visual reprojection residuals can be referred to T. Qin et al. (2018).

As demonstrated in Equation (11), the accuracy of the solution heavily relies on the quality of feature extraction and subsequent tracking. However, current feature extraction and tracking methods are highly sensitive to illumination changes, which are common and inevitable in urban waterway scenarios, such as when a USV navigates beneath poorly illuminated bridges. To address this issue, we employ a lightweight convolutional neural network (CNN) proposed in Lin et al. (2024) for robust feature extraction. This network is capable of processing raw images on a CPU within a few milliseconds, generating an illumination-invariant feature map. This significantly enhances the robustness of both feature extraction and the subsequent Kanade–Lucas–Tomasi tracking method, ensuring reliable performance even under varying lighting conditions.

This least-square problem can be efficiently solved using the Ceres Solver (Agarwal et al. 2023). Once solved, the estimation of the most recent state is passed to the LIS as an initial guess for LiDAR scan-to-map registration.

3.2.3 | Factor Graph Optimization

Given the relative motion constraints from LIS odometry, loop-closure constraints, IMU preintegration constraints, a factor graph problem for consistent state estimation of MAP at the pose level can be formulated as

$$\arg \min_{\boldsymbol{x}} \left\{ \|\mathbf{r}_0\|_{\Sigma_0}^2 + \sum_{i=0}^{n-1} \left(\|\mathbf{r}_i(\tilde{\mathcal{Z}}_{\mathcal{L}_i}, \boldsymbol{x})\|_{\Sigma_{\mathcal{L}_i}}^2 + \|\mathbf{r}_L(\tilde{\mathcal{Z}}_{\mathcal{L}_i}, \boldsymbol{x})\|_{\Sigma_{\mathcal{L}_i}}^2 \right) + \sum \|\mathbf{r}_L(\tilde{\mathcal{Z}}_{\mathcal{L}_{pq}}, \boldsymbol{x})\|_{\Sigma_{\mathcal{L}_{pq}}}^2 \right\}, \quad (12)$$

where \mathbf{r}_0 represents the residual error on the initial state; $\tilde{\mathcal{Z}}_{\mathcal{L}_i}$ is the residual of IMU preintegration measurements between LiDAR keyframes i and $i+1$; $\tilde{\mathcal{Z}}_{\mathcal{L}_i}$ is the residual of the relative LiDAR odometry transformation between consecutive LiDAR keyframes i and $i+1$; $\tilde{\mathcal{Z}}_{\mathcal{L}_{pq}}$ is the residual of the relative transformation between nonconsecutive LiDAR keyframes p and q , obtained through loop-closure detection. $\Sigma_{\mathcal{L}_{pq}}$ is the covariance matrix determined by the fitness score of ICP matching (Besl and McKay 1992).

In scenarios involving long-term USV navigation, accumulated drift errors cannot be fully mitigated through loop-closure detection, as the USV may not revisit historical scenes. To address this, it is essential to incorporate absolute references, such as GNSS measurements, into the factor graph formulation. This helps correct drift and maintain global consistency over extended periods. Given the longitude-latitude-altitude coordinates $(\lambda_{\mathcal{L}_i}, \varphi_{\mathcal{L}_i}, h_{\mathcal{L}_i})$ of the LiDAR frame \mathcal{L}_i , we can first

convert it to the Earth-Centered, Earth-Fixed (ECEF) coordinate system:

$$\begin{cases} \text{ECEF } x_{\mathcal{L}_i} = (N + h_{\mathcal{L}_i}) \cos \varphi_{\mathcal{L}_i} \cos \lambda_{\mathcal{L}_i}, \\ \text{ECEF } y_{\mathcal{L}_i} = (N + h_{\mathcal{L}_i}) \cos \varphi_{\mathcal{L}_i} \sin \lambda_{\mathcal{L}_i}, \\ \text{ECEF } z_{\mathcal{L}_i} = [N(1 - e^2) + h_{\mathcal{L}_i}] \sin \varphi_{\mathcal{L}_i}, \end{cases} \quad (13)$$

where $N = a / \sqrt{1 - e^2 \sin^2 \varphi_{\mathcal{L}_i}}$, and $e^2 = (a^2 - b^2)/a^2$; a and b are the long semiaxis and short semiaxis of the reference ellipsoid corresponding to the ECEF coordinate system, respectively. The coordinate origin of the local ENU frame \mathcal{W} in the ECEF coordinate system, denoted as $(\text{ECEF } x_{\mathcal{L}_i}, \text{ECEF } y_{\mathcal{L}_i}, \text{ECEF } z_{\mathcal{L}_i})$, can also be calculated in the same way. Then the conversion from the ECEF coordinate system to the local ENU frame can be formulated as

$${}^W \mathbf{p}_{\text{LLA}} = {}^W \mathbf{R}_{\text{ECEF}} \begin{bmatrix} \text{ECEF } x_{\mathcal{L}_i} - \text{ECEF } x_{\mathcal{W}} \\ \text{ECEF } y_{\mathcal{L}_i} - \text{ECEF } y_{\mathcal{W}} \\ \text{ECEF } z_{\mathcal{L}_i} - \text{ECEF } z_{\mathcal{W}} \end{bmatrix}, \quad (14a)$$

$${}^W \mathbf{R}_{\text{ECEF}} = \begin{bmatrix} -\sin \lambda_{\mathcal{W}} & \cos \lambda_{\mathcal{W}} & 0 \\ -\sin \varphi_{\mathcal{W}} \cos \lambda_{\mathcal{W}} & -\sin \varphi_{\mathcal{W}} \sin \lambda_{\mathcal{W}} & \cos \varphi_{\mathcal{W}} \\ \cos \varphi_{\mathcal{W}} \cos \lambda_{\mathcal{W}} & \cos \varphi_{\mathcal{W}} \sin \lambda_{\mathcal{W}} & \sin \varphi_{\mathcal{W}} \end{bmatrix}, \quad (14b)$$

where $\lambda_{\mathcal{W}}, \varphi_{\mathcal{W}}$ are the longitude and latitude of the origin of \mathcal{W} , respectively.

The residual of the GNSS measurement on the state of keyframe i , namely, the GNSS factor, can be calculated as

$$\mathbf{r}_{\text{gnss}}(\tilde{\mathcal{Z}}_{\text{LLA}}, \boldsymbol{x}_i) = {}^W \mathbf{p}_{\text{LLA}} - {}^W \mathbf{p}_{\mathcal{L}_i}. \quad (15)$$

It is important to note that the pose of the first LiDAR frame is coarsely initialized using the orientation from the nearest IMU measurement. As a result, the local point cloud map is not rigorously aligned with the local ENU frame until GNSS measurements are incorporated. This alignment with the ENU frame is essential for global consistency. Once the poses of the LiDAR keyframes are corrected through the first factor graph optimization that includes GNSS measurements, the incremental point cloud map must be rebuilt to ensure accurate scan-to-map registration. To save the computational burden of frequent map rebuilding, factor graph optimization is triggered only when a sufficient batch of GNSS measurements has been accumulated. This approach balances computational efficiency with the need for accurate global alignment.

3.3 | Predictive Kinematic-Compliant Collision Avoidance

3.3.1 | Dynamic Obstacle Modeling and Prediction

In this study, the distribution of static and dynamic obstacles is discretized using a 2D occupancy grid map centered on and

moving with the USV. Each grid cell is assigned two key attributes: an occupancy probability indicating the likelihood of occupancy by an obstacle, and a 2D velocity vector representing the motion state of the obstacle within the cell. To robustly estimate the occupancy probability and velocity state of each grid from noisy LiDAR observations, we introduce the concept of a random finite set (RFS) (Ronald 2007). In an RFS, the number and the states of the elements are random but finite. In this subsection, let X denote an RFS and $\mathbf{x}^{(i)} = [p_x, p_y, v_x, v_y]^T \in \mathbb{R}^4$ denote the state of an element in X . Thus, X can be expressed as

$$X = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(i)}, \dots, \mathbf{x}^{(N)}\}. \quad (16)$$

Here, $N \in \mathbb{N}$ is a random variable representing the number of elements in X , also referred to as the cardinality of X . Note that $\mathbf{x}^{(i)}$ is defined in the coordinate system of \mathcal{W} . Using this concept, each occupied grid cell can be mathematically represented by an element of X . This representation enables the dynamic occupancy grid map to be formulated as a multiobject-tracking problem within a Bayesian framework (Mahler 2003).

3.3.1.1 | SMC-PHD Filter. In this part, we introduce the realization of the multiobject Bayes filter based on a tractable technique SMC (Vo et al. 2005) and multiobject density approximation utilizing probability hypothesis density (PHD) (Mahler 2003), together with the process procedure of how the dynamic occupancies are recursively updating. Within this framework, the multitarget states are modeled as an RFS, where each particle represents a potential target state hypothesis with an associated weight. After completing the PHD prediction and update steps, the spatial intensity of the PHD function is approximated by aggregating particle weights within discrete grid cells. The occupancy probability of each cell is then derived to enable probabilistic mapping of complex obstacle distributions.

PHD approximates the full multiobject density using the first statistical moment. It is proved in Mahler (2003) that PHD is the best-fit approximation of multiobject density in an information-theoretic sense. According to the definition, we get

$$D_X(\mathbf{x}) = E \left[\sum_{\mathbf{x}^{(i)} \in X} \delta(\mathbf{x} - \mathbf{x}^{(i)}) \right], \quad (17)$$

where $E[\cdot]$ denotes the expectation while $\delta(\cdot)$ the Dirac function. Since $D_X(\mathbf{x})$ is an intensity distribution, this leads to another important property:

$$\int D_X(\mathbf{x}) d\mathbf{x} = E[|X|], \quad (18)$$

where $|X|$ is the cardinality of the RFS $|X|$ and corresponds to the expected number of objects in the map. This property lays the fundamental for occupancy status estimation. Based hereon, the multiobject Bayes filter can be rederived in PHD representation. Then the PHD of an RFS is propagated in the prediction and the update step by direct operation on the particles, which is known as SMC-PHD filter. The main procedures of the SMC-PHD filter

contain prediction, update, particle birth, and resampling, all of which are directly operated on the particles with different weights.

Suppose the posterior PHD of the RFS X_k at time k is already known as $D_{X_k}(\mathbf{x}_k)$. In the prediction step, the RFS $X_{k+1|k}$ can be treated as the union of two subsets, namely, $X_{k+1|k} = S_{k+1|k} \cup B_{k+1|k}$, where $S_{k+1|k}$ represents the survival objects from X_k , and $B_{k+1|k}$ is the newly born objects. Then the prior PHD of the RFS $X_{k+1|k}$ is derived as

$$\begin{aligned} D_{X_{k+1|k}}(\mathbf{x}_{k+1}) &= \gamma_{b,k+1|k}(\mathbf{x}_{k+1}) + P_s \int \pi_{k+1|k}(\mathbf{x}_{k+1}|\mathbf{x}_k) \cdot D_{X_k}(\mathbf{x}_k) d\mathbf{x}_k \\ &\approx \sum_{\tilde{\mathbf{x}}_{b,k+1|k}^{(i)} \in B_{k+1|k}} \omega_{b,k+1|k} \cdot \delta(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}_{b,k+1|k}^{(i)}) \\ &\quad + P_s \left[\sum_{\tilde{\mathbf{x}}_{s,k+1|k}^{(i)} \in S_{k+1|k}} \omega_{s,k}^{(i)} \cdot \delta(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}_{s,k+1|k}^{(i)}) \right]. \end{aligned} \quad (19)$$

The subset $B_{k+1|k}$ is modeled as Poisson Point Process with the prior intensity $\gamma_{b,k+1|k}(\mathbf{x}_{k+1})$, and the corresponding new-born particles are generated around a new observation $\mathbf{z}_{k+1}^{(i)} \in Z_k$ at time k sharing the same weight calculated as

$$\omega_{b,k+1|k} = \frac{\int \gamma_{b,k+1|k}(\mathbf{x}_{k+1}) d\mathbf{x}_{k+1}}{M_K \cdot L_{b,k+1|k}}, \quad (20)$$

where $\int \gamma_{b,k+1|k}(\mathbf{x}_{k+1}) d\mathbf{x}_{k+1}$ is the user-defined parameter that controls the expected number of new-born objects, M_K is the number of elements in Z_k , and $L_{b,k+1|k}$ is the number of new-born particles for a measurement point. To avoid redundancy of massive particle operations, the new-born particles generate only in the grid cells with new observations. Their positions are sampled around measurement points with the Gaussian noise model, while the velocity of each particle is sampled through an initial velocity estimation method based on fast Euclidean Clustering (Cao et al. 2022) and Kuhn–Munkres matching method to accelerate the convergence speed (Chen et al. 2024).

The other subset $S_{k+1|k}$ is modeled as Multi-Bernoulli Mixture Process with the survival probability P_s . Note that the weight $\omega_{s,k}^{(i)}$ of a survival particle is unchanged while the state is propagated by a constant velocity model with random propagating noise. The state transition $\pi_{k+1|k}(\mathbf{x}_{k+1}|\mathbf{x}_k)$ can be formulated as

$$\tilde{\mathbf{x}}_{s,k+1|k}^{(i)} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tilde{\mathbf{x}}_{s,k}^{(i)} + \boldsymbol{\mu}, \quad (21)$$

where $\boldsymbol{\mu} \in \mathbb{R}^4$ is a noise vector subjected to a normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{Q} \in \mathbb{R}^{4 \times 4})$.

In update step, the posterior PHD of X_{k+1} is given by

$$D_{X_{k+1}}(\mathbf{x}_{k+1}) \approx \sum_{\tilde{\mathbf{x}}_{k+1}^{(i)} \in X_{k+1}} \omega_{k+1}^{(i)} \cdot \delta(\mathbf{x}_{k+1} - \tilde{\mathbf{x}}_{k+1}^{(i)}), \quad (22)$$

where the particle state $\tilde{\mathbf{x}}_{k+1}^{(i)}$ remains unchanged from the prediction step and the weight $\omega_{k+1}^{(i)}$ is updated as follows:

$$\omega_{k+1}^{(i)} \approx \omega_{k+1|k}^{(i)} \cdot \left[1 - P_d + \sum_{\mathbf{z}_{k+1} \in Z_{k+1}} \frac{P_d g_{k+1}(\mathbf{z}_{k+1} | \tilde{\mathbf{x}}_{k+1}^{(i)})}{P_{cl}(\mathbf{z}_{k+1}) + C_{k+1}(\mathbf{z}_{k+1})} \right], \quad (23a)$$

$$C_{k+1}(\mathbf{z}_{k+1}) = \sum_{\tilde{\mathbf{x}}_{k+1|k}^{(i)} \in X_{k+1}} P_d \cdot \omega_{k+1|k}^{(i)} \cdot g_{k+1}(\mathbf{z}_{k+1} | \tilde{\mathbf{x}}_{k+1}^{(i)}), \quad (23b)$$

where $g_{k+1}(\mathbf{z}_{k+1} | \tilde{\mathbf{x}}_{k+1}^{(i)})$ is a measurement likelihood function subjected to a normal distribution, and $P_{cl}(\mathbf{z}_{k+1})$ is the clutter intensity function. For further details, we refer readers to Chen et al. (2024) and Nuss et al. (2018).

Therefore, with the combination of the PHD filter and SMC implementation, the multitarget Bayes filter is more computationally efficient in the state estimation of dynamic occupancy grid map and intuitive in a physical sense. In Figure 4, the SMC-PHD filter algorithm flowchart is illustrated.

3.3.1.2 | Dynamic Occupancy Grid Prediction. First, based on the property that the cardinality expectation of the RFS can be utilized to estimate the occupancy status in a proximity space, we can calculate the occupancy probability of a single grid cell with

$$pr_{occ,k}^{(j)} = \int D_{X_k^{(j)}}(\mathbf{x}) d \approx \sum_{\tilde{\mathbf{x}}_k^{(i)} \in X_k^{(j)}} \omega_k^{(i)}, \quad (24)$$

where j represents the index of grid cell in the current robot centered map, $X_k^{(j)}$ represents the subset of the RFS X_k corresponding to the grid cell j . The grid cell j would be deemed occupied if the $pr_{occ,k}^{(j)}$ exceeds a certain threshold.

The 2D velocity of an occupancy grid cell can be calculated from the statistical moment. If grid cell j contains a certain number of particles, the mean velocity component in 1D direction can be approximated by

$$\bar{v}_{\mu,k}^{(j)} \approx \frac{1}{pr_{occ,k}^{(j)}} \cdot \sum_{\tilde{\mathbf{x}}_k^{(i)} \in X_k^{(j)}} \omega_k^{(i)} \cdot v_{\mu,k}^{(i)}, \quad \mu \in \{x, y\}. \quad (25)$$

To anticipate future occupancy information for predictive motion planning, we implement a fast ray tracing method

(Amanatides and Woo 1987) to implement the forward prediction. This method efficiently integrates anticipated trajectories of dynamic obstacles over a short-term horizon into the occupancy grid map. The resultant layers of the occupancy grid map that incorporate short-term predictive information regarding collision risk will be transmitted to the path planning aspects for the subsequent decision-making process.

3.3.2 | Kinematic-Compliant Path Planning

To tackle dynamic obstacle avoidance in urban waterway environments, we develop a hybrid 2D path planning framework integrating the front-end kinematic-compliant path searching method based on A^* , and B-spline trajectory optimization. Within the front-end searching, primitive generation and heuristic search are performed based on the USV kinematics. The back-end trajectory optimization further guarantees the satisfaction of the feasibility, smoothness and safety constraints.

3.3.2.1 | Motion Primitive Generation. According to the differential flatness of the USV dynamic proved in Huang et al. (2023), the characteristic of USV motion can be described by a linear time-invariant (LTI) system. Thus, the motion primitives for the node expansion in A^* can be generated by the following state-space equation:

$$\dot{\mathbf{x}}(t) \approx \mathbf{Ax}(t) + \mathbf{Bu}(t),$$

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{0}_{2 \times 2} & \mathbf{I}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_{2 \times 2} \end{bmatrix} \mathbf{x}(t) = [\mathbf{p}(t)^T, \dot{\mathbf{p}}(t)^T]^T, \quad \mathbf{u}(t) = \ddot{\mathbf{p}}(t), \end{aligned} \quad (26)$$

where $\mathbf{x}(t) \in \mathbb{R}^4$ is the state vector, $\mathbf{p}(t) \in \mathbb{R}^2$ is the plane position in global-fixed coordinate frame, and acceleration $\ddot{\mathbf{p}}(t) = [a_x(t), a_y(t)]^T$ is defined as control input $\mathbf{u}(t)$. Given the current state \mathbf{x}_c and a set of evenly distributed discrete control inputs:

$$\left\{ -\mathbf{u}_{\max}, -\frac{n-1}{n}\mathbf{u}_{\max}, \dots, \frac{n-1}{n}\mathbf{u}_{\max}, \mathbf{u}_{\max} \right\}, \quad (27)$$

where the acceleration boundaries are $[-\mathbf{u}_{\max}, \mathbf{u}_{\max}]$. We can obtain $(2n+1)^3$ primitives by the Euler integration method with a time interval.

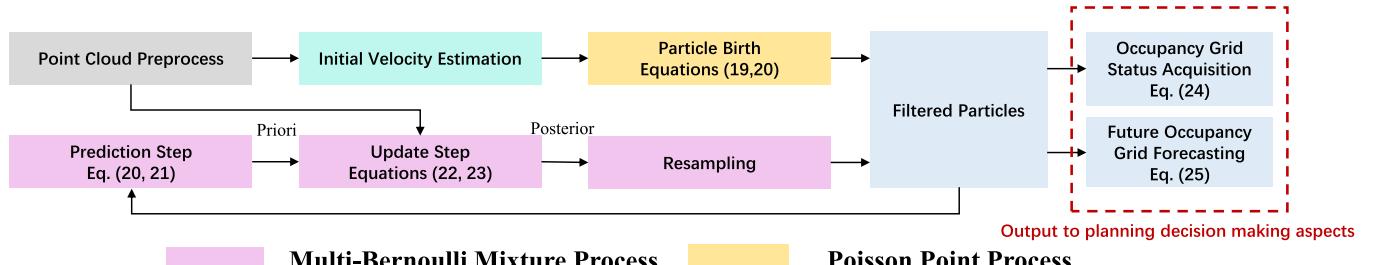


FIGURE 4 | Flowchart of SMC-PHD filter algorithm. [Color figure can be viewed at wileyonlinelibrary.com]

Due to the limited turning ability of USVs, the curvature radius of primitives should be considered. The formula for calculating the corresponding curvature radius is

$$R(t) = \frac{\|\dot{\mathbf{p}}\|^3}{|\dot{p}_x \ddot{p}_y - \ddot{p}_x \dot{p}_y|}. \quad (28)$$

During the expansion, the primitives with a curvature radius smaller than the predefined R_{\min} will be neglected.

3.3.2.2 | Heuristically Searching. A heuristic is used to quantify the cost from the current searching node to the goal. A well-designed heuristic plays an important role in guiding the path smoothly and efficiently to the goal. Therefore, we employ a cost function that is based on time and control input cost, which is defined as (Zhou et al. 2019)

$$J(T) = \int_0^T (\|\mathbf{u}(t)\|^2 + \varepsilon) dt, \quad (29)$$

where ε denotes the weight of time. With the characteristic of the LTI system, the USV trajectory can be represented by two independent time-parameterized cubic polynomial functions. After applying Pontryagin Minimum Principle, Equation (29) has a closed-form solution without iteration given the current node state \mathbf{x}_c and the goal state \mathbf{x}_g :

$$p_\mu^* = \frac{1}{6}\alpha_\mu t^3 + \frac{1}{2}\beta_\mu t^2 + v_{\mu c} T + p_{\mu c}, \quad \mu \in \{x, y\}, \quad (30a)$$

$$\begin{bmatrix} \alpha_\mu \\ \beta_\mu \end{bmatrix} = \frac{1}{T^3} \begin{bmatrix} -12 & 6T \\ 6T & -2T^2 \end{bmatrix} \begin{bmatrix} p_{\mu g} - p_{\mu c} - v_{\mu c} T \\ v_{\mu g} - v_{\mu c} \end{bmatrix}, \quad (30b)$$

$$J_{cg}^*(T) = \sum_{\mu \in \{x, y\}} \frac{1}{3}\alpha_\mu^2 T^3 + \frac{1}{2}\alpha_\mu \beta_\mu T^2 + \beta_\mu^2 T, \quad (30c)$$

where $p_{\mu g}$, $v_{\mu g}$, $p_{\mu c}$, and $v_{\mu c}$ are the current and goal positions and velocities. The optimal duration T^* can be obtained through $\partial J_{cg}(T)/\partial T = 0$ after substituting Equation (30b) into Equation (30c). The heuristic h_c is defined by $J_{cg}(T^*)$.

The actual cost of an optimal path consisting of K primitives from the start state \mathbf{x}_s to the current state \mathbf{x}_c can be calculated with

$$g_c = \sum_{k=1}^K (\|\mathbf{u}_k\|^2 + \varepsilon)\tau. \quad (31)$$

Then we calculate the node cost through $f_c = g_c + h_c$.

3.3.2.3 | Trajectory Optimization. The trajectory generated in Section 3.3.2(a) is coarse and may be in the vicinity of obstacles. Therefore, the initially generated trajectory should be further optimized. using uniform B-spline parametrization for further optimization. The trajectory is first parametrized with uniform B-spline control points (K. Qin 1998). This yields the advantage of sparsity and alleviates the computation burden of

optimization. Moreover, due to the convex hull property of B-splines, the trajectory will be safe as long as the control points are collision free. A cubic B-spline defined by $N+1$ control points $\{\mathbf{C}_j \in \mathbb{R}^2 | j = 0, 1, \dots, N\}$ is utilized to parametrized the initial trajectory. The subset $\{\mathbf{C}_j \in \mathbb{R}^2 | j = 3, \dots, N-3\}$ is employed to construct the convex optimization problem to obtain the optimal trajectory.

This optimization incorporates gradient information from the Euclidean distance field and dynamic constraints, constructing smoothness, safety distance, and dynamic feasibility costs as quadratic soft constraints. The cost in total is then defined as

$$f_{\text{total}} = \lambda_s f_s + \lambda_c f_c + \lambda_d f_d, \quad (32)$$

where f_s, f_c, f_d are the costs associated with smoothness, collision, and dynamic constraints, and $\lambda_s, \lambda_c, \lambda_d$ are the corresponding weights that trade off smoothness, safety, and feasibility.

The smoothness cost f_s is defined as an elastic band cost function to capture the geometric information of the trajectory:

$$f_s = \sum_{i=2}^{N-2} \|\mathbf{C}_{i+1} + \mathbf{C}_{i-1} - 2\mathbf{C}_i\|^2. \quad (33)$$

To make the trajectory of the collision part tend to keep away from obstacles, the collision safety cost f_c is formulated as

$$f_c = \begin{cases} \sum_{i=3}^{N-3} (d(\mathbf{C}_i) - d_{\min})^2, & d(\mathbf{C}_i) \leq d_{\min} \\ 0, & d(\mathbf{C}_i) > d_{\min}, \end{cases} \quad (34)$$

where $d(\mathbf{C}_i)$ is the distance between the i th control point and the nearest obstacle and d_{\min} is the predefined safety distance.

To avoid exceeding the dynamics of the USV, constraints on speed and acceleration are incorporated into the cost function f_d , which is formulated as

$$f_d = \sum_{\mu \in \{x, y\}} \left\{ \sum_{j=2}^{N-2} [(v_{\mu j} - v_{\max})^2 + (a_{\mu j} - a_{\max})^2] \right\}. \quad (35)$$

This optimization converges rapidly with the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (a quasi-Newton optimization algorithm) proposed by D. C. Liu and Nocedal (1989).

4 | Experimental Results and Discussion

4.1 | Hardware and Software Setup

Our technique verification USV is a typical catamaran equipped with dual propellers. Detailed specifications are listed in Table 1. As shown in Figure 5, the overall system architecture for the USV can be divided into three distinct parts: onboard sensors, onboard processor, and ground control terminal. The onboard sensors are responsible for

perception data acquisition for the autonomous navigation framework, including a mechanical LiDAR, a monocular industrial camera, and an IMU with a GNSS antenna. They have been compactly integrated within a 3D printed bracket (located at the bottom right in Figure 5). For more details about sensor specifications, readers are referred to Table 2. The onboard processor part consists of a high-performance onboard computer with a compact volume to implement computation tasks of measurement processing, state estimation and motion planning, and an advanced flight control unit (FCU) to convert velocity and angular velocity commands into pulse-width modulation signals for the dual differential propulsions. To facilitate remote control and monitoring of the onboard processor, a pair of wireless network bridges establishes a local area network between the onboard computer and the laptop. All the related software is developed with C++ in a Linux environment (Ubuntu 20.04) with open-source Robot Operating System Noetic.

TABLE 1 | Technical specification of USV.

Items	Contents
Propulsion	Dual 60 N propellers (T80-60-ALA)
Onboard sensors	3D LiDAR, Camera, IMU, GPS receiver
Onboard computer	Nvidia Jetson Orin NX
Flight control unit	CUAV X7 + pro
Wireless network	UIISP airMAX Rocket 5AC Lite

Abbreviations: GPS, global positioning system; IMU, inertial measurement unit; LiDAR, light detection and ranging; UIISP, ubiquiti internet service provider; USV, unmanned surface vehicle.

4.2 | SLAM Experiments

To assess the enhancement achieved by the LiDAR-camera–IMU-coupled system, we conducted an experiment for accuracy assessment on a lake situated within a campus. The Xsens-MTi 680-dk IMU was paired with a ZED F9P GNSS receiver module. Throughout data collection, the ZED-F9P receiver received real-time Radio Technical Commission for Maritime Services (standard for GNSS correction data) streams to obtain ground truth RTK solutions. Adequate sky visibility was ensured during the LiDAR, camera, and IMU data collection process. The USV was manually maneuvered at an average speed of 1.0 m/s. Throughout the experiment, the reported horizontal covariance of the RTK output constantly remained 0.000196 m². Note that the filtered RTK localization result was not integrated into our system to maintain fairness. Both the modified and original systems operated with the same sensor combinations and parameters, such as key pose adding distance and voxel resolution for sampling down planar points (0.2 m). Upon re-running on the recorded data set, the trajectories plotted for both the modified and original systems are depicted in Figure 6, alongside the RTK ground truth trajectory. The modified system demonstrates decimeter-level accuracy, meeting up with the demands of autonomous USV navigation. Figure 6 illustrates the trajectory of the modified system closely overlapping the RTK ground truth trajectory, even upon close inspection in three detailed regions. Contrarily, the trajectory of the original system exhibits noticeable deviation from the ground truth, despite implementing Umeyama alignment with the RTK trajectory. Quantitative analysis of absolute and relative position errors using the Evaluation of Odometry (a toolkit for visual odometry/SLAM system assessment) toolbox (Grupp 2017), as shown in Table 3, underscores the substantial improvement achieved.

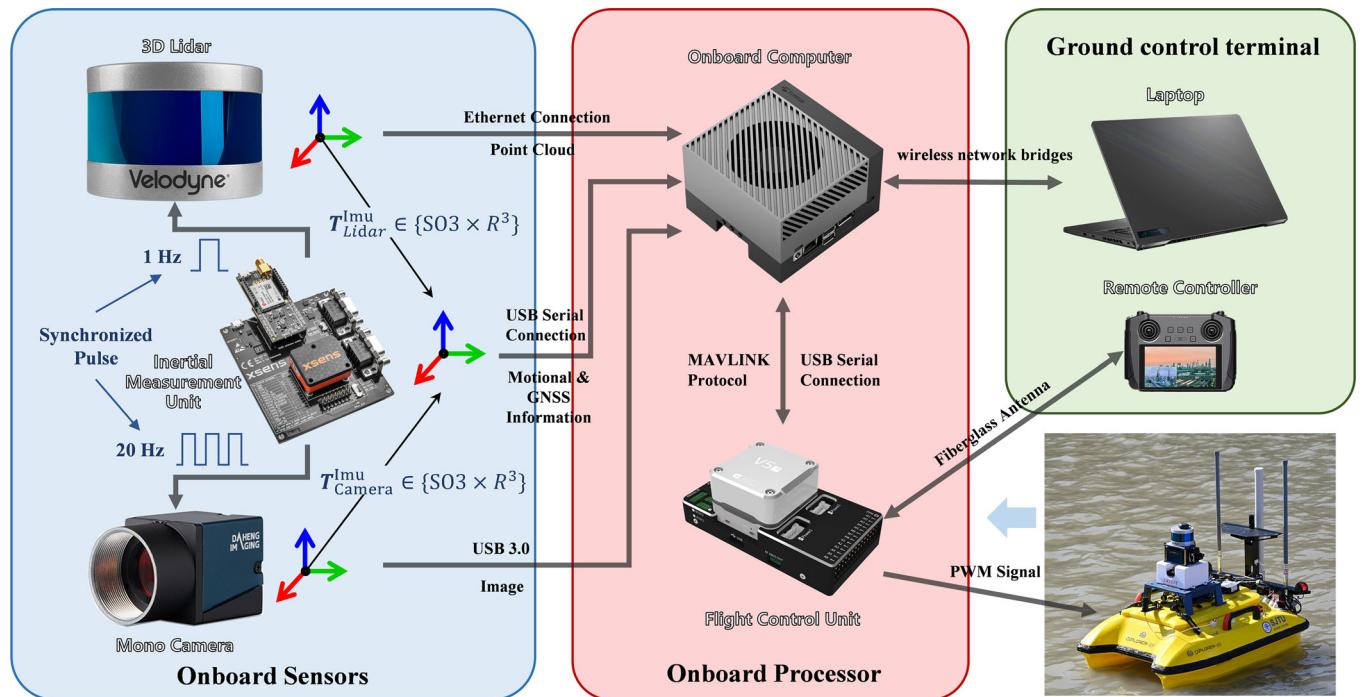


FIGURE 5 | Overview of hardware system architecture and electrical connections. 3D, three-dimensional; GNSS, Global Navigation Satellite System; IMU, inertial measurement unit; MAVLINK, microair vehicle link; PWM, pulse-width modulation; USB, universal serial bus. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 | Sensor specification for hardware setup.

<i>LiDAR VLP-16 Puck</i>	
Vertical/horizontal resolution	$2^\circ \times 0.1^\circ$
Vertical field of view	(-15.0° , 15.0°)
Horizontal field of view	(-180.0° , 180°)
Sample rate	10 Hz
Range/range accuracy	100 m/ ± 3 cm
<i>Camera MER-301-125U3C</i>	
Resolution	2048×1536
Sensor type	1/8" Global shutter CMOS
Pixel format	Bayer RG8, Bayer RG10
Sample rate	20 Hz
Interface	I/O USB3.0
<i>Lens</i>	
Focal length	3.5 mm fixed
Aperture	F/1.4–16
Vertical/horizontal field of view	$86.7^\circ/117.7^\circ$
Minimum focus distance	200 mm
<i>IMU Xsens MTi 680-dk (RTK/GNSS receiver)</i>	
Roll/Pitch RMS	0.2°
Yaw RMS	0.5°
Gyroscope in-run bias	$8^\circ/\text{h}$
Accelerometer bias	$10(x,y)$ $15(z)\mu\text{g}$
Magnetometer RMS	1 mG

Abbreviations: CMOS, complementary metal oxide semiconductor; GNSS, Global Navigation Satellite System; I/O, input/output; IMU, inertial measurement unit; LiDAR, light detection and ranging; RMS, root mean square; RTK, real-time kinematics.

To further illustrate the improvement, two screencasts of the mapping and localization visualization process were captured as the USV approaches the bridge in the middle of the lake and compared in Figure 7. The fuchsia trajectories correspond to the direct IMU propagation result between the scan-to-map optimization. The factor graph optimization updated the IMU bias estimation based on the LiDAR odometry factor. However, if the LiDAR odometry estimation deteriorates and becomes inconsistent with the IMU propagation result, the system will prioritize the LiDAR odometry factor, resulting in a zigzag IMU propagation trajectory. In contrast, the modified system, which employed the advanced mapping data structure iKD-Tree, benefited from improved efficiency and matching accuracy in the scan-to-map registration. Consequently, the IMU propagation trajectory tended to be smoother and closer to the LiDAR odometry trajectory, as drawn by the cyan trajectory in Figure 7. Additionally, the mapping of the bridge structure in the modified system was clearer due to the consistent estimation between IMU and LiDAR odometry.

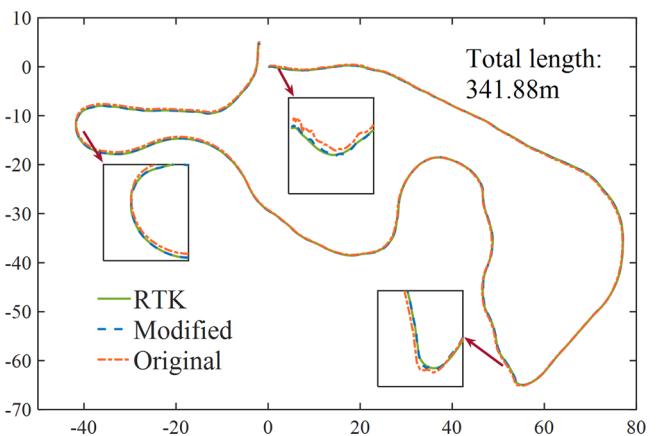


FIGURE 6 | Trajectories of the modified and original localization subsystems compared with RTK ground truth for the accuracy experiment. RTK, real-time kinematics. [Color figure can be viewed at wileyonlinelibrary.com]

4.3 | Global Navigation Experiments

A more rigorous loop-closure waterway experiment was conducted to assess the practicability, long-term robustness, and stability of the proposed autonomous navigation system. In this experiment, no prior static obstacle map is provided; instead, only a series of predefined reachable waypoints were sent to the onboard processor. The system periodically replanned trajectories between adjacent waypoints in real-time based on the most recent LiDAR measurement data. Throughout the autonomous navigation process, measurements from LiDAR, camera, IMU, and GNSS were seamlessly integrated for localization and mapping. To ensure safety, a hand-held remote controller was available as a backup for emergency interventions. However, challenges remained, particularly when navigating through long bridge tunnels, where unsighted risks posed potential hazards to the USV. Figure 8a illustrates the overall USV trajectory of the global navigation experiment, which covers a loop-closure waterway, spanning a total distance of 3.6 km in a clockwise direction. Throughout the experiment, the USV navigated autonomously at an average speed of 1.0 m/s without any collisions or human intervention. The USV successfully reached each predefined waypoint as intended, demonstrating the system's ability to achieve consistent global localization without significant drift. This outcome highlights the robustness and reliability of the proposed navigation system in real-world conditions.

To further evaluate the performance of the proposed system, the absolute trajectory error (ATE) of the LiDAR-camera-IMU-GNSS localization trajectories, compared with the RTK solution, is plotted in Figure 9 for quantitative analysis. Additionally, the time series of the RTK covariance squared root, reported by the ZED-F9P, is plotted in Figure 9 to distinguish between valid and invalid RTK solutions. Under clear sky conditions, the RTK status remained stable, with a corresponding covariance of 0.000196 m^2 . However, when the USV sailed through bridge tunnels, the reported covariance increased significantly, indicating high uncertainty in RTK localization. A total of 12 peaks were observed in the RTK covariance squared root time series,

TABLE 3 | Quantitative comparison of the localization subsystem before and after modification.

Translation error compared with ground truth (unit: meter)				
	Mean	Maximum	Minimum	RMSE
Modified	0.0952 (0.028%)	0.5014 (0.146%)	0.0010	0.1093
Original	0.6399 (0.187%)	2.3820 (0.697%)	0.0138	0.7748

Abbreviation: RMSE, root mean square error.

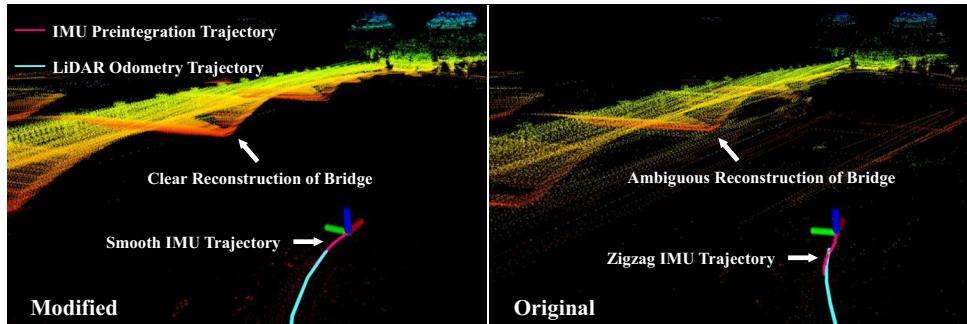


FIGURE 7 | Snapshots of localization and mapping visualization. IMU, inertial measurement unit; LiDAR, light detection and ranging; RTK, real-time kinematics. [Color figure can be viewed at wileyonlinelibrary.com]

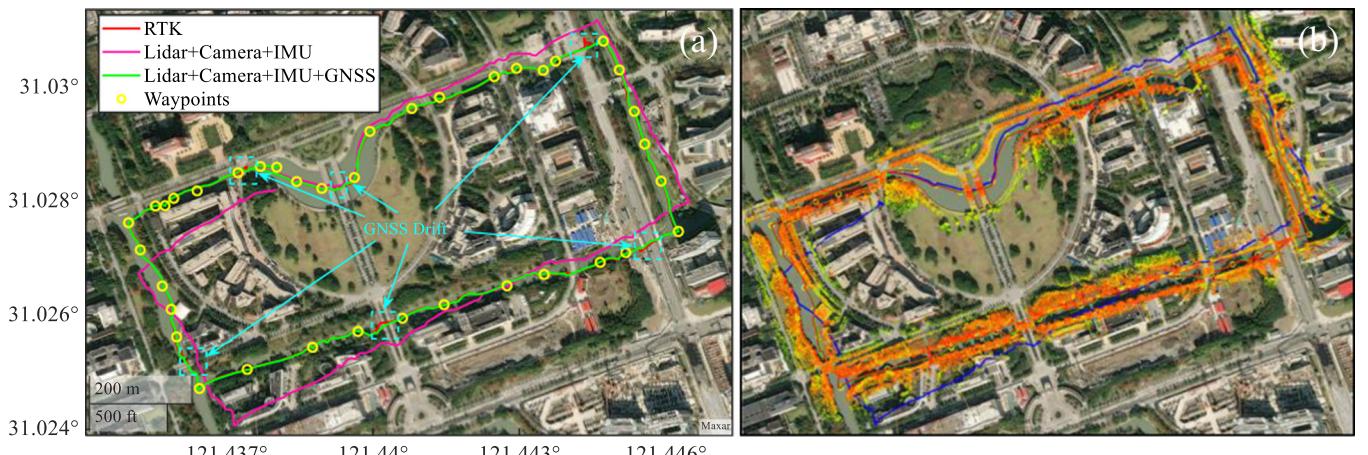


FIGURE 8 | Long-term loop-closure navigation results. (a) Diagram of the long-term loop-closure navigation experiment results: RTK, LiDAR-camera-IMU and LiDAR-camera-IMU-GNSS trajectories and (b) diagram of overall mapping result of LiDAR upon satellite image. GNSS, Global Navigation Satellite System; IMU, inertial measurement unit; LiDAR, light detection and ranging; RTK, real-time kinematics. [Color figure can be viewed at wileyonlinelibrary.com]

corresponding precisely to the 12 bridge tunnels encountered during the long-term navigation experiment. Since the RTK solution under these conditions is unreliable as a ground truth reference, the statistical analysis of the ATE excludes the invalid RTK data. According to the statistics presented in Table 4, the mean ATE of the LiDAR-camera-IMU-GNSS localization was 0.8172 m, with a 0.7577-m standard deviation, meeting the requirements for USV autonomous navigation. A top-view screencast of the global mapping result, scaled to align with a satellite image, is shown in Figure 8b, demonstrating excellent consistency and alignment with the satellite imagery. Most importantly, the enhanced LiDAR-camera-IMU localization system enabled the USV to maintain stable and continuous state estimation, which is critical for trajectory planning and tracking control, particularly in GNSS-attenuated areas, such as bridge tunnels.

However, the maximum ATE value peaked at 4.0472 m at the 1500-s mark in Figure 9, corresponding to the USV's passage through a long bridge tunnel. The interior surfaces of this tunnel lacked sufficient longitudinal geometric features for extraction from LiDAR scans, causing a LiDAR-slip phenomenon along the long corridor. This limitation caused a degradation in LiDAR odometry, leading to noticeable localization drift after exiting the tunnel. Figure 10 illustrates the entire drifting process and highlights that the LiDAR-Visual-IMU odometry was successfully corrected once the GNSS signal became available again. This correction underscores the system's ability to recover from drift and maintain reliable localization in challenging environments.

To demonstrate the critical role of GNSS in long-term navigation tasks, an ablation study was conducted by replaying the

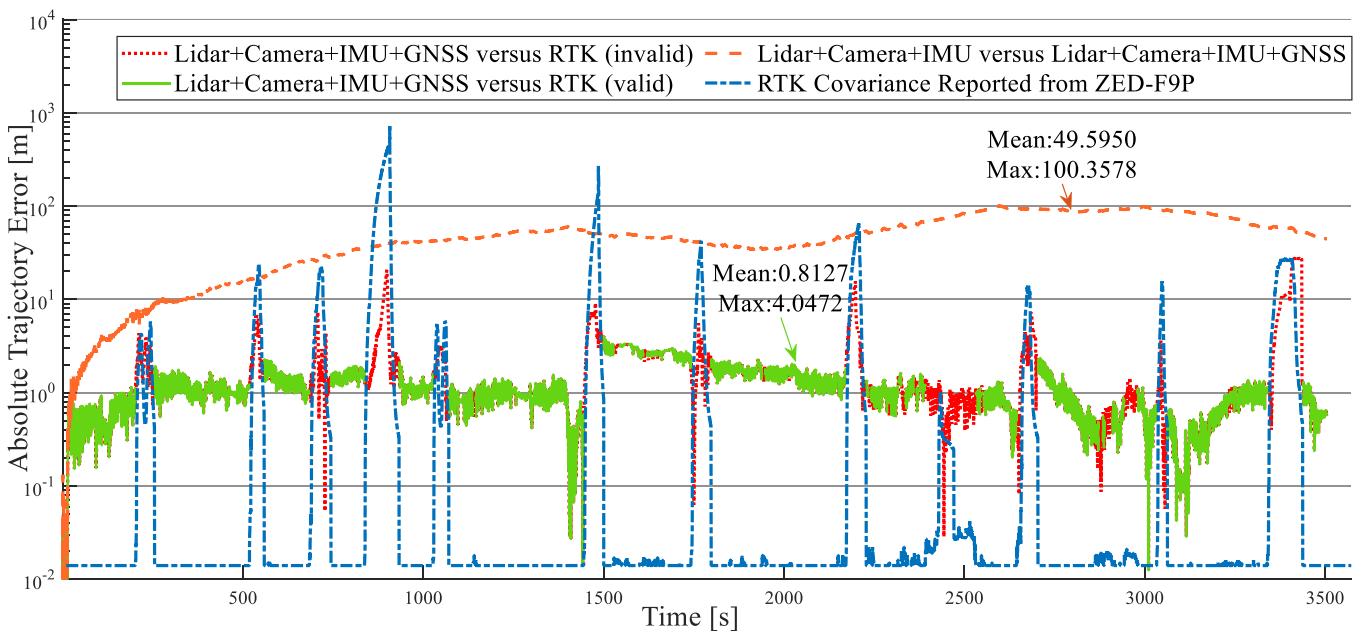


FIGURE 9 | Absolute trajectory error of the localization trajectories versus RTK solution with respect to the traveled duration under long-term navigation experiment. GNSS, Global Navigation Satellite System; IMU, inertial measurement unit; LiDAR, light detection and ranging; RTK, real-time kinematics. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 4 | Absolute trajectory error comparison between localization trajectories with/without GNSS correction.

	Max ATE (m)	Mean ATE (m)
LiDAR + Camera + IMU + GNSS	4.0742	0.8127
LiDAR + Camera + IMU	100.3578	49.5950

Abbreviations: ATE, absolute trajectory error; GNSS, Global Navigation Satellite System; IMU, inertial measurement unit; LiDAR, light detection and ranging.

collected data set to obtain localization results using only LiDAR, camera, and IMU. As shown in Figure 8a, a significant localization error was observed in the fuchsia pose trajectory, which drifted away from the waterways onto the ground. Without GNSS correction, the maximum ATE reached 100.3578 m, accounting for 2.78% of the total traveled distance, while the mean ATE was 49.5950 m, as depicted by the right dotted curve in Figure 9. The drifting error accumulated as the traveled distance increased. Several factors contributed to this drift in the absence of GNSS or loop-closure detection: reflection of water waves, which distorted sensor measurements, sparse LiDAR measurements over the broad water surface, and lack of planar and corner features for robust feature matching. Without GNSS correction, the long-term navigation mission would inevitably fail to follow the predefined waypoints accurately, as illustrated by the deviation from the satellite map. This highlights the indispensable role of GNSS in maintaining global consistency and providing an absolute localization reference to correct drift, especially when the USV travels long distances without revisiting historical scenes for extended periods. The integration of GNSS measurements enables correcting historical keyframe poses, thus ensuring robust and accurate long-term navigation performance.

4.4 | Dynamic Avoidance Experiments

As the global navigation experiment has already demonstrated the system's capability for static obstacle avoidance, this section focuses on evaluating the dynamic obstacle-avoidance performance of the proposed predictive kinematic-compliant planner. Field experiments were conducted using two USVs in four distinct scenarios:

1. The Own USV navigates toward the Obstacle USV in a head-on configuration.
2. The Obstacle USV crosses in front of the Own USV from its starboard side to the port side.
3. The Obstacle USV aggressively overtakes the Own USV.
4. The Own USV follows a zig-zag path defined by five waypoints, while the Obstacle USV autonomously tracks a similar zig-zag path, intentionally creating four collision avoidance events.

The operational setup for the two USVs is detailed as follows. The Own USV employs the proposed collision avoidance algorithm at a constant speed of 1 m/s without remote controller intervention, except for the manual setting of goal points. Localization is provided by a multisensor fusion framework. Figure 11 illustrates the logical process of predictive collision avoidance in the presence of a dynamic obstacle, supported by three snapshots from a real-world implementation. Then, in the first three scenarios, the Obstacle USV is manually controlled to simulate collision risks. In the fourth scenario, the Obstacle USV operates in auto mode, utilizing its built-in autopilot functionality to track mission waypoints at a cruise speed of 1 m/s. Notably, the two USVs operate independently without any communication between them.

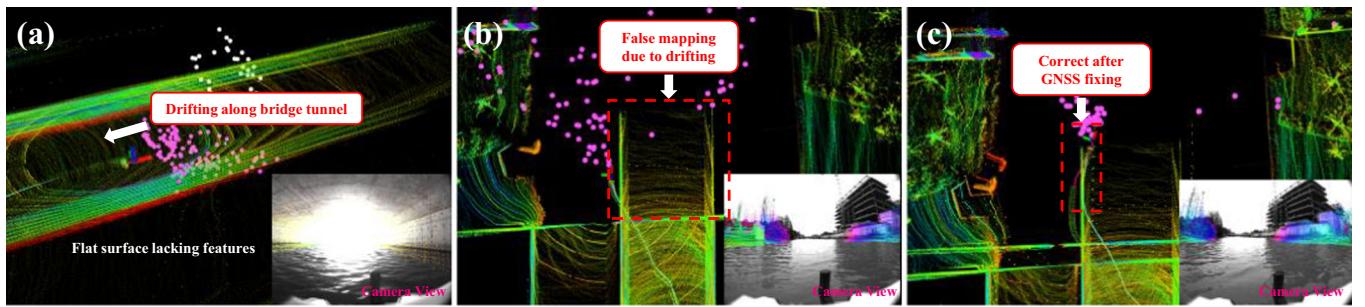


FIGURE 10 | Process of LiDAR-Visual-IMU odometry drifting and correction after GNSS fixing. (a) “LiDAR-slip” degeneration occurred in a long bridge tunnel, (b) false mapping after getting out of the bridge, and (c) poses were corrected immediately after GNSS available. GNSS, Global Navigation Satellite System; IMU, inertial measurement unit; LiDAR, light detection and ranging. [Color figure can be viewed at wileyonlinelibrary.com]

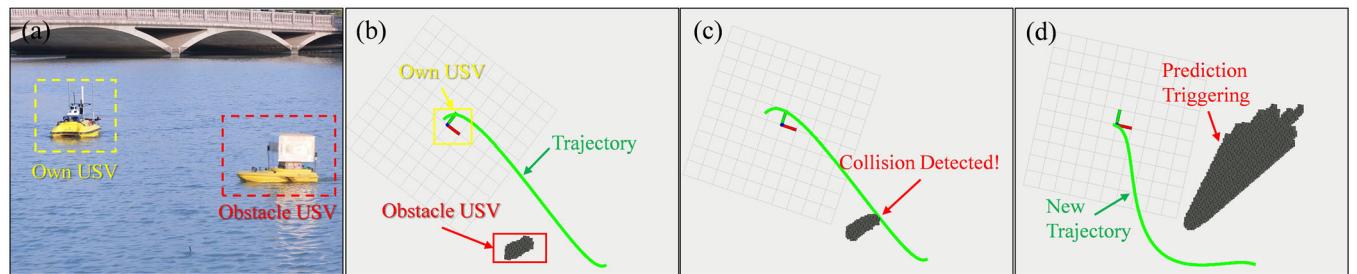


FIGURE 11 | Snapshots of the predictive collision avoidance process confronted with a dynamic obstacle. (a) Real scene of field experiment, (b) Own USV follows the original trajectory before collision risk is detected, (c) Obstacle USV steps on the original trajectory, and (d) a new natural safer trajectory is replanned after triggering the prediction of velocity grids via ray casting. [Color figure can be viewed at wileyonlinelibrary.com]

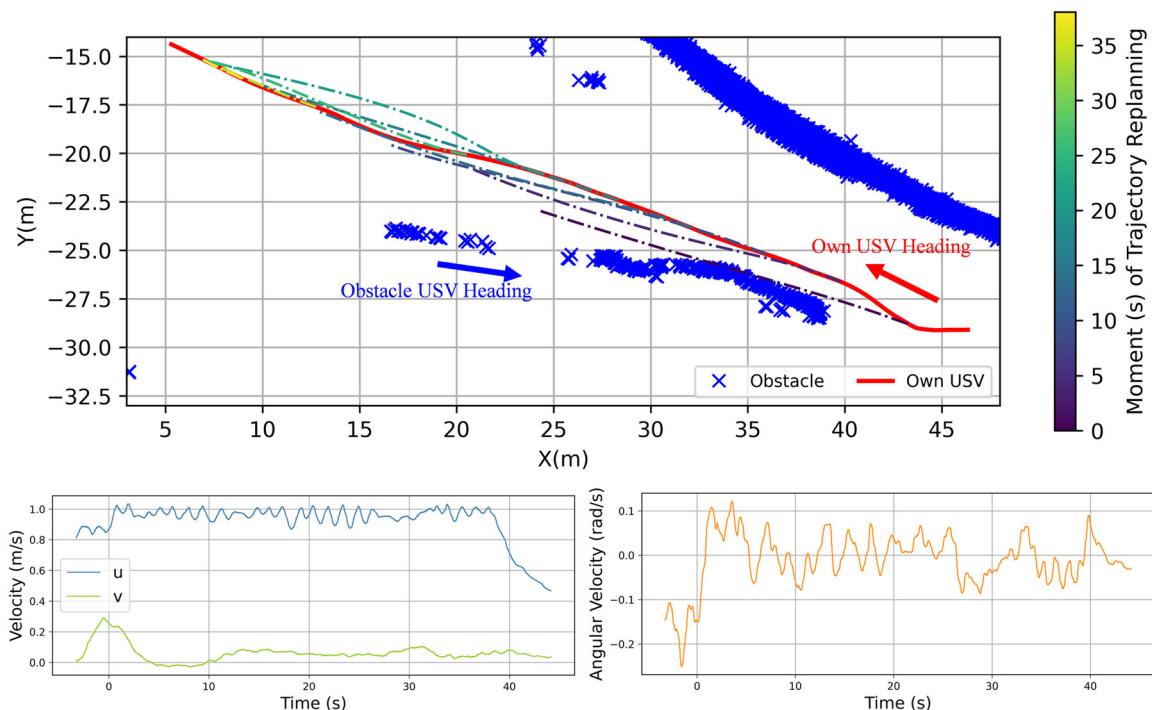


FIGURE 12 | Process of the head-on scenario: Actual and replanned trajectories of the Own USV; distribution of the obstacles; surge, sway, and yaw velocities plotted at the bottom. USV, unmanned surface vehicle. [Color figure can be viewed at wileyonlinelibrary.com]

The results of the field experiments for the four distinct scenarios are presented in Figures 12–15. In all experiments, obstacles (represented by blue forks) were accurately identified by the SMC-PHD-based dynamic occupancy grid and

successfully avoided by the kinematic-compliant planner. For clarity, the predicted occupancy grid is not displayed in the figures. The historical positions of the Obstacle USV are indicated by the traces of blue forks, with blue arrows

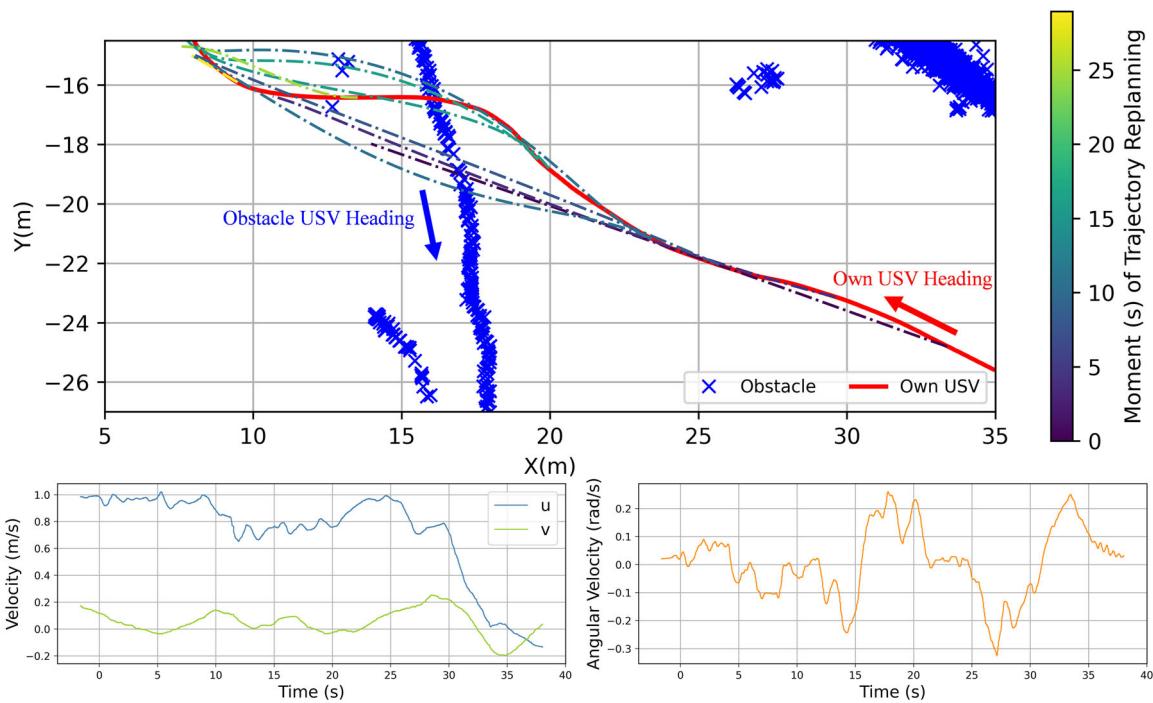


FIGURE 13 | Process of the crossing scenario: Actual and replanned trajectories of the Own USV; distribution of the obstacles; surge, sway, and yaw velocities plotted at the bottom. USV, unmanned surface vehicle. [Color figure can be viewed at wileyonlinelibrary.com]

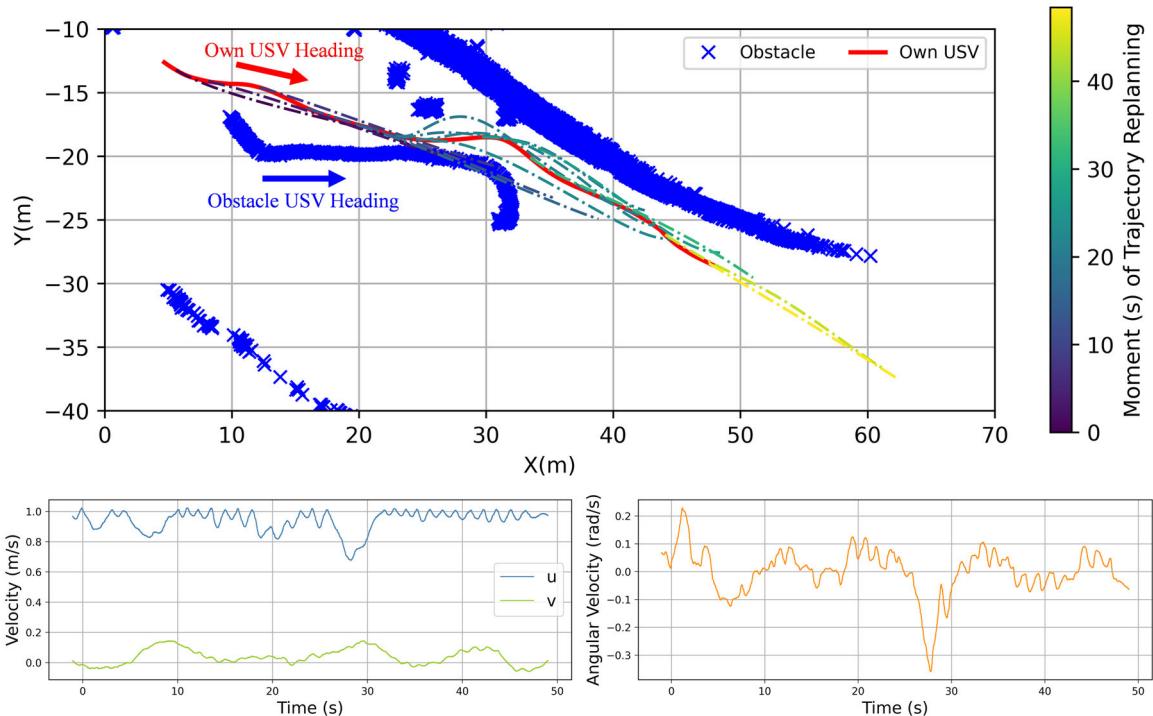


FIGURE 14 | Process of the Overtaking Scenario: Actual and replanned trajectories of the Own USV; distribution of the obstacles; surge, sway, and yaw velocities plotted at the bottom. USV, unmanned surface vehicle. [Color figure can be viewed at wileyonlinelibrary.com]

denoting their corresponding headings, while the Own USV's trajectory is represented by a red solid curve. The sequences of repeatedly replanned trajectories are depicted as dot-dash lines, with varying colors indicating the timing of replanning. A color bar on the left side of each figure illustrates the replanning

sequence, where darker colors correspond to earlier replanning moments.

In Figure 12, the upper subplot depicts the head-on scenario, where the Own USV navigated smoothly toward its goal at a

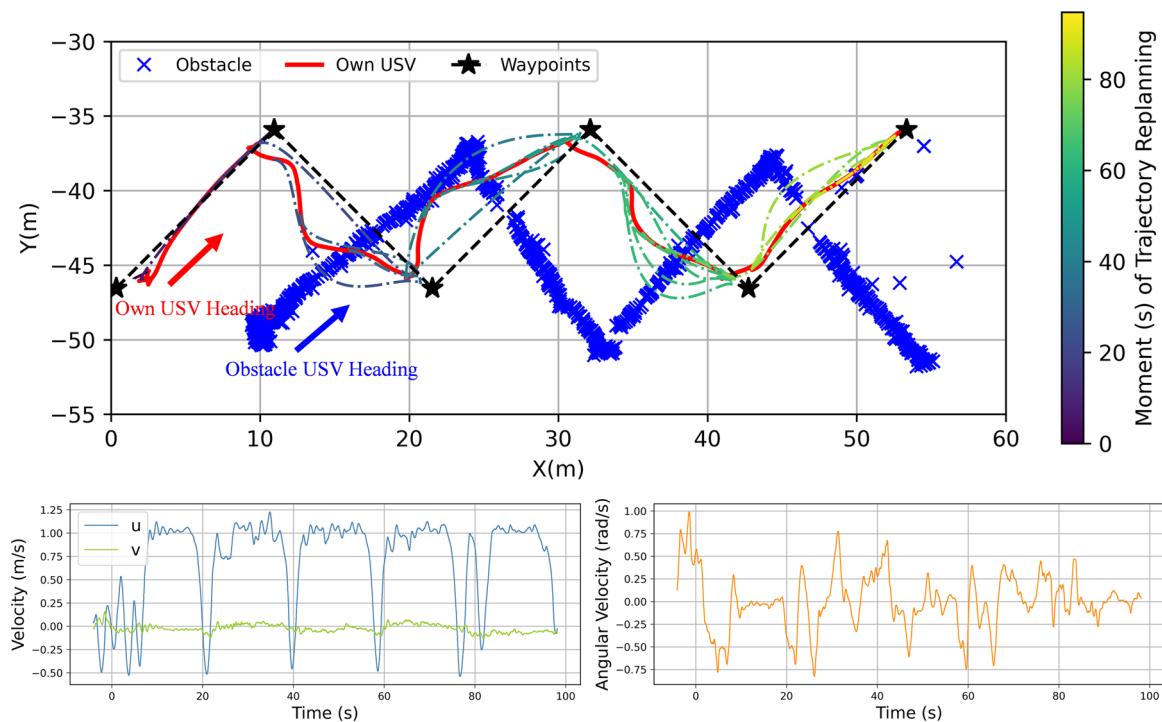


FIGURE 15 | Process of the zig-zag maneuvering scenario: Actual and replanned trajectories of the Own USV; distribution of the obstacles; surge, sway, and yaw velocities plotted at the bottom. USV, unmanned surface vehicle. [Color figure can be viewed at wileyonlinelibrary.com]

speed of approximately 1 m/s, as shown in the bottom-left subplot. Since the trajectories of the two USVs did not overlap as the Obstacle USV approached, the prediction mechanism was not triggered. Instead, the Own USV was guided to its starboard side through periodic replanning.

In contrast, the crossing scenario, illustrated in Figure 13, demonstrates the significant influence of the prediction mechanism. When the Obstacle USV intersected the nearly straight trajectory replanned at approximately 10 s, a collision risk was detected, triggering the forward prediction. From 10 to 20 s, the trajectories were repeatedly replanned, guiding the Own USV to turn toward its starboard side, effectively maneuvering behind the Obstacle USV. This process was also reflected in the linear and angular velocities: the Own USV initially maintained a speed of around 1 m/s before decelerating and turning with a negative angular velocity. The avoidance behavior was notably conservative due to the inflation distance of the occupancy grid being set to 0.6 m and the safety distance in the B-spline trajectory optimization set to 2.0 m. Despite this conservatism, the overall trajectory of the Own USV remained smooth, demonstrating the effectiveness of the trajectory optimization process.

Figure 14 illustrates the overtaking scenario, where the Obstacle USV approached the Own USV from the starboard side, intentionally posing a collision threat. Leveraging the prediction of the dynamic obstacle, the Own USV initially adjusted its trajectory with a pronounced turn and gradually returned to its original course as the collision risk diminished. Throughout the process, the Own USV maintained a safe distance from both the dynamic obstacle and the lake's border. In summary, the prediction of the Obstacle USV's potentially hazardous behavior enabled the Own USV to navigate conservatively yet safely.

The most complex scenario, involving zig-zag maneuvering, is depicted in Figure 15. Despite four collision avoidance events during the process, the Own USV successfully tracked all five waypoints without collisions, even under successive collision threats. This achievement was attributed to the prediction of the Obstacle USV's trajectory, which allowed all reference trajectories to be replanned behind the Obstacle USV. This capability significantly enhanced the safety of USV navigation in dynamic environments and demonstrated the robustness of the proposed predictive kinematic-compliant planner. However, it is worth noting that the Own USV's actual trajectory deviated from the straight lines connecting adjacent waypoints, indicating a tendency toward conservative avoidance behavior.

5 | Conclusion

In this study, we developed an enhanced autonomous navigation system for USVs, demonstrating significant potential for application in GNSS-attenuated and dynamically crowded urban waterways. Through extensive real-world experiments, the system exhibited exceptional performance in robust low-drift localization, long-term autonomous navigation, and efficient dynamic obstacle avoidance. A globally consistent, low-drift, and robust state estimation framework was achieved by integrating LiDAR, camera, IMU, and GNSS measurements within a factor graph optimization framework. This approach proved particularly effective in GNSS-challenging urban waterways. Additionally, the system's capability to autonomously navigate unknown environments without prior maps was successfully demonstrated in a loop-closure waterway experiment. Furthermore, predictive motion planning for dynamic obstacle avoidance was efficiently realized by incorporating an SMC-PHD filter-based dynamic occupancy grid map into a grid-based local reactive planner.

Our future work will pursue the following three directions. First, we will develop a more sophisticated multisensor fusion architecture with tightly coupling raw GNSS measurements (e.g., pseudo-range and Doppler measurements), adaptive sensor fusion, and smooth switching strategies, to cover more complex scenarios of urban waterway. Second, we will merge the state-of-the-art machine learning method, like deep CNN with self-attention mechanism, to advance the current dynamic occupancy map framework, offering more accurate prediction of dynamic obstacles. Third, we will add compliance with COLREGS traffic rules into our collision avoidance planner, and implement more deep-in theoretical investigation and comparison to other state-of-the-art methods in a high-fidelity simulation environment.

Acknowledgments

This study was funded by National Natural Science Foundation of China (No. 52271284) and the Oceanic Interdisciplinary Program of Shanghai Jiao Tong University (No. SL2021ZD201).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Agarwal, S., and K. Mierle, The-Ceres-Solver-Team. 2023. Ceres Solver. <https://github.com/ceres-solver/ceres-solver>.
- Amanatides, J., and A. Woo. 1987. "A Fast Voxel Traversal Algorithm for Ray Tracing." In *Proceedings of the EuroGraphics*, edited by G. Maréchal, Vol. 87. Elsevier Science Publishing Company. https://www.researchgate.net/publication/2611491_A_Fast_Voxel_Traversal_Algorithm_for_Ray_Tracing.
- Besl, P. J., and N. D. McKay. 1992. "A Method for Registration of 3-D Shapes." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14, no. 2: 239–256. <https://doi.org/10.1109/34.121791>.
- Cai, Y., W. Xu, and F. Zhang. 2021. "iKd-Tree: An Incremental K-D Tree for Robotic Applications." arXiv Preprint, arXiv:2102.10808. <https://arxiv.org/abs/2102.10808>.
- Cao, Y., Y. Wang, Y. Xue, H. Zhang, and Y. Lao. 2022. "FEC: Fast Euclidean Clustering for Point Cloud Segmentation." *Drones* 6, no. 11: 325. <https://doi.org/10.3390/drones6110325>.
- Chen, G., W. Dong, P. Peng, J. Alonso-Mora, and X. Zhu. 2024. "Continuous Occupancy Mapping in Dynamic Environments Using Particles." *IEEE Transactions on Robotics* 40: 64–84. <https://doi.org/10.1109/TRO.2023.3323841>.
- Cheng, Y., M. Jiang, J. Zhu, and Y. Liu. 2021. "Are We Ready for Unmanned Surface Vehicles in Inland Waterways? The Usvinland Multisensor Dataset and Benchmark." *IEEE Robotics and Automation Letters* 6, no. 2: 3964–3970. <https://doi.org/10.1109/LRA.2021.3067271>.
- Christian, F., C. Luca, D. Frank, and S. Davide. 2015. "IMU Pre-integration on Manifold for Efficient Visual-Inertial Maximum-a-Posteriori Estimation." Paper Presented at the Robotics: Science and Systems.
- Cole, B., and G. Schamberg. 2022. "Unscented Kalman Filter for Long-Distance Vessel Tracking in Geodetic Coordinates." *Applied Ocean Research* 124: 103205. <https://doi.org/10.1016/j.apor.2022.103205>.
- Dellaert, F., and M. Kaess. 2017. "Factor Graphs for Robot Perception." *Foundations and Trends in Robotics* 6: 1–139. <https://doi.org/10.1561/2300000043>.
- Friedman, J. H., J. L. Bentley, and R. A. Finkel. 1977. "An Algorithm for Finding Best Matches in Logarithmic Expected Time." *ACM Transactions on Mathematical Software* 3: 209–226. <https://doi.org/10.1145/355744.355745>.
- Fun Sang Cepeda, M., M. S. Freitas Machado, F. H. Sousa Barbosa, et al. 2023. "Exploring Autonomous and Remotely Operated Vehicles in Offshore Structure Inspections." *Journal of Marine Science and Engineering* 11, no. 11: 2172.
- Gonzalez-Garcia, A., I. Collado-Gonzalez, R. Cuan-Urquiza, C. Sotelo, D. Sotelo, and H. Castañeda. 2022. "Path-Following and LiDAR-Based Obstacle Avoidance via NMPC for an Autonomous Surface Vehicle." *Ocean Engineering* 266: 112900. <https://doi.org/10.1016/j.oceaneng.2022.112900>.
- Grupp, M. 2017. evo: Python Package for the Evaluation of Odometry and SLAM. <https://github.com/MichaelGrupp/evo>.
- Han, S., L. Wang, and Y. Wang. 2022. "A COLREGs-Compliant Guidance Strategy for an Underactuated Unmanned Surface Vehicle Combining Potential Field With Grid Map." *Ocean Engineering* 255: 111355. <https://doi.org/10.1016/j.oceaneng.2022.111355>.
- Huang, T., Z. Xue, Z. Chen, and Y. Liu. 2023. "Efficient Trajectory Planning and Control for USV With Vessel Dynamics and Differential Flatness." Paper Presented at the 2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM).
- Li, Y., R. Wang, D. Gao, and Z. Liu. 2023. "A Floating-Waste-Detection Method for Unmanned Surface Vehicle Based on Feature Fusion and Enhancement." *Journal of Marine Science and Engineering* 11, no. 12: 2234. <https://doi.org/10.3390/jmse11122234>.
- Li, Y., and H. Zhang. 2022. "Collision Avoidance Decision Method for Unmanned Surface Vehicle Based on an Improved Velocity Obstacle Algorithm." *Journal of Marine Science and Engineering* 10, no. 8: 1047. <https://doi.org/10.3390/jmse10081047>.
- Lin, Y., S. Wang, Y. Jiang, and B. Han. 2024. "Breaking of Brightness Consistency in Optical Flow With a Lightweight CNN Network." *IEEE Robotics and Automation Letters* 9, no. 8: 6840–6847. <https://doi.org/10.1109/LRA.2024.3414250>.
- Liu, D., X. Gao, and C. Huo. 2022. "Motion Planning for Unmanned Surface Vehicle Based on a Maneuverability Mathematical Model." *Ocean Engineering* 265: 112507. <https://doi.org/10.1016/j.oceaneng.2022.112507>.
- Liu, X., Z. Hu, Z. Sun, J. Lu, W. Xie, and W. Zhang. 2023. "A VIO-Based Localization Approach in GPS-Denied Environments for an Unmanned Surface Vehicle." Paper presented at the 2023 International Conference on Advanced Robotics and Mechatronics (ICARM).
- Liu, W., Y. Liu, and R. Bucknall. 2019. "A Robust Localization Method for Unmanned Surface Vehicle (USV) Navigation Using Fuzzy Adaptive Kalman Filtering." *IEEE Access* 7: 46071–46083. <https://doi.org/10.1109/ACCESS.2019.2909151>.
- Liu, W., Y. Liu, and R. Bucknall. 2023. "Filtering Based Multi-Sensor Data Fusion Algorithm for a Reliable Unmanned Surface Vehicle Navigation." *Journal of Marine Engineering & Technology* 22, no. 2: 67–83. <https://doi.org/10.1080/20464177.2022.2031558>.
- Liu, Z., X. Liu, and F. Zhang. 2023. "Efficient and Consistent Bundle Adjustment on Lidar Point Clouds." *IEEE Transactions on Robotics* 39, no. 6: 4366–4386. <https://doi.org/10.1109/TRO.2023.3311671>.
- Liu, D. C., and J. Nocedal. 1989. "On the Limited Memory BFGS Method for Large Scale Optimization." *Mathematical Programming* 45, no. 1: 503–528. <https://doi.org/10.1007/BF01589116>.
- Mahler, R. P. S. 2003. "Multitarget Bayes Filtering via First-Order Multi-target Moments." *IEEE Transactions on Aerospace and Electronic Systems* 39, no. 4: 1152–1178. <https://doi.org/10.1109/TAES.2003.1261119>.
- Nuss, D., S. Reuter, M. Thom, et al. 2018. "A Random Finite Set Approach for Dynamic Occupancy Grid Maps With Real-Time

- Application." *International Journal of Robotics Research* 37, no. 8: 841–866. <https://doi.org/10.1177/0278364918775523>.
- Qin, K. 1998. "General Matrix Representations for B-Splines." Paper presented at the Proceedings of the Pacific Graphics '98. Sixth Pacific Conference on Computer Graphics and Applications (Cat. No. 98EX208).
- Qin, T., P. Li, and S. Shen. 2018. "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator." *IEEE Transactions on Robotics* 34, no. 4: 1004–1020. <https://doi.org/10.1109/TRO.2018.2853729>.
- Ramkumar Sudha, S. K., D. Mishra, and I. A. Hameed. 2024. "A Coverage Path Planning Approach for Environmental Monitoring Using an Unmanned Surface Vehicle." *Ocean Engineering* 310: 118645. <https://doi.org/10.1016/j.oceaneng.2024.118645>.
- Ronald, M. 2007. *Statistical Multisource-Multitarget Information Fusion*. Artech.
- Shan, T., E. Brendan, R. Carlo, and R. Daniela. 2021. "LVI-SAM: Tightly-Coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping." In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 5692–5698. IEEE. <https://doi.org/10.1109/ICRA48506.2021.9561996>.
- Shan, T., B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus. 2020. "LIO-SAM: Tightly-Coupled Lidar Inertial Odometry via Smoothing and Mapping." In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5135–5142. IEEE. <https://doi.org/10.1109/IROS45743.2020.9341176>.
- Shan, T., W. Wang, B. Englot, C. Ratti, and D. Rus. 2020. "A Receding Horizon Multi-Objective Planner for Autonomous Surface Vehicles in Urban Waterways." In *2020 59th IEEE Conference on Decision and Control (CDC)*, 4085–4092. <https://doi.org/10.1109/CDC42340.2020.9304298>.
- Thoms, A., G. Earle, N. Charron, and S. Narasimhan. 2023. "Tightly Coupled, Graph-Based DVL/IMU Fusion and Decoupled Mapping for SLAM-Centric Maritime Infrastructure Inspection." *IEEE Journal of Oceanic Engineering* 48, no. 3: 663–676. <https://doi.org/10.1109/JOE.2023.3265742>.
- Tuna, T., J. Nubert, Y. Nava, S. Khattak, and M. Hutter. 2024. "X-ICP: Localizability-Aware LiDAR Registration for Robust Localization in Extreme Environments." *IEEE Transactions on Robotics* 40: 452–471. <https://doi.org/10.1109/TRO.2023.3335691>.
- Vo, B.-N., S. Singh, and A. Boucet. 2005. "Sequential Monte Carlo Methods for Multitarget Filtering With Random Finite Sets." *IEEE Transactions on Aerospace and Electronic Systems* 41, no. 4: 1224–1245. <https://doi.org/10.1109/TAES.2005.1561884>.
- Volden, Ø., A. Stahl, and T. I. Fossen. 2022. "Vision-Based Positioning System for Auto-Docking of Unmanned Surface Vehicles (USVs)." *International Journal of Intelligent Robotics and Applications* 6, no. 1: 86–103. <https://doi.org/10.1007/s41315-021-00193-0>.
- Volden, Ø., D. Cabecinhas, A. Pascoal, and T. I. Fossen. 2023. "Development and Experimental Evaluation of Visual-Acoustic Navigation for Safe Maneuvering of Unmanned Surface Vehicles in Harbor and Waterway Areas." *Ocean Engineering* 280: 114675. <https://doi.org/10.1016/j.oceaneng.2023.114675>.
- Wang, W., D. Fernández-Gutiérrez, R. Doornbusch, et al. 2023. "Roboat III: An Autonomous Surface Vessel for Urban Transportation." *Journal of Field Robotics* 40, no. 8: 1996–2009. <https://doi.org/10.1002/rob.22237>.
- Wang, W., B. Gheneti, L. A. Mateos, F. Duarte, C. Ratti, and D. Rus. 2019. "Roboat: An Autonomous Surface Vehicle for Urban Waterways." In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 6340–6347. IEEE. <https://doi.org/10.1109/IROS40897.2019.8968131>.
- Wang, Y., C. Liu, J. Liu, et al. 2024. "A Laser-Based SLAM Algorithm of the Unmanned Surface Vehicle for Accurate Localization and Mapping in an Inland Waterway Scenario." *Journal of Marine Science and Engineering* 12, no. 12: 2311. <https://doi.org/10.3390/jmse12122311>.
- Wang, H., Y. Yin, Q. Jing, Z. Cao, Z. Shao, and D. Guo. 2024. "Berthing Assistance System for Autonomous Surface Vehicles Based on 3D LiDAR." *Ocean Engineering* 291: 116444. <https://doi.org/10.1016/j.oceaneng.2023.116444>.
- Wen, G., J. Fu, H. Lu, J. Sun, and H. Shen. 2024. "Robust Collision Avoidance and Path-Following of USVs With Reduced Conservativeness: A Control Barrier Function-Based Approach." *Journal of Field Robotics* 42, no. 4: 1388–1400. <https://doi.org/10.1002/rob.22456>.
- Xiang, M., H. Chai, X. Yin, Z. Du, and K. Jin. 2023. "Precise Navigation of USV Based on PPP-RTK/MEMS in the Offshore Environment." *Marine Geodesy* 46, no. 5: 441–459. <https://doi.org/10.1080/01490419.2023.2223764>.
- Zhang, J., and S. Singh. 2017. "Low-Drift and Real-Time Lidar Odometry and Mapping." *Autonomous Robots* 41: 401–416. <https://doi.org/10.1007/s10514-016-9548-2>.
- Zhang, J., and S. Singh. 2018. "Laser-Visual-Inertial Odometry and Mapping With High Robustness and Low Drift." *Journal of Field Robotics* 35: 1242–1264. <https://doi.org/10.1002/rob.21809>.
- Zhao, L., Y. Bai, and J. K. Paik. 2024. "Optimal Coverage Path Planning for USV-Assisted Coastal Bathymetric Survey: Models, Solutions, and Lake Trials." *Ocean Engineering* 296: 116921. <https://doi.org/10.1016/j.oceaneng.2024.116921>.
- Zhou, B., F. Gao, L. Wang, C. Liu, and S. Shen. 2019. "Robust and Efficient Quadrotor Trajectory Generation for Fast Autonomous Flight." *IEEE Robotics and Automation Letters* 4: 3529–3536. <https://doi.org/10.1109/LRA.2019.2927938>.
- Zhou, B., Y. He, K. Qian, X. Ma, and X. Li. 2021. "S4-SLAM: A Real-Time 3D LiDAR SLAM System for Ground/Watersurface Multi-Scene Outdoor Applications." *Autonomous Robots* 45: 77–98. <https://doi.org/10.1007/s10514-020-09948-3>.
- Zhu, H., and Y. Ding. 2023. "Optimized Dynamic Collision Avoidance Algorithm for USV Path Planning." *Sensors* 23, no. 9: 4567. <https://doi.org/10.3390/s23094567>.