Model-based Boat Recognition for Urban River Navigation using Waterborne LiDAR and Scan Matching

Kazuki O.[1*], Tetsu Y.[1], Masafumi N.[1], Nobuaki K.[2], Etsuro S.[2]

[1]*Shibaura Institute of Technology, 3-7-5, Toyosu, Koto-ku, Tokyo 135-8548, Japan*

[2]*Tokyo University of Marine Science and Technology, 2-1-6 Etchujima, Koto-ku, Tokyo 135-8533, Japan*

ah20091@shibaura-it.ac.jp

Abstract: *This study aimed to investigate the development of autonomous boating technology, with a focus on improving the automatic detection of surrounding boats and collision avoidance functions. Despite their importance, these functions remain underdeveloped. More than 40% of boating accidents in 2023 were caused by collisions. For small boats in particular, reliance on manual observation by crew members when monitoring their surroundings is a major challenge. This study proposes a new boat detection methodology that uses model-based object recognition light detection and ranging (LiDAR) and scan matching installed on boats. This methodology uses the real-time measurement capabilities of LiDAR for multi-purpose mapping. Conventional methodologies rely on image processing technologies such as global navigation satellite system (GNSS)-based position sharing and deep learning models such as faster region-based convolutional neural networks. However, GNSS position sharing is not feasible in non-GNSS positioning environments such as under bridges. Furthermore, image processing methodologies that use deep learning still have issues because they depend on pretrained models and require training data of boats facing different directions. On the other hand, this study proposes a method for recognizing boats in urban rivers using LiDAR. First, scan matching is performed using fast point feature histograms and point-to-plane iterative closest point to align point clouds. Next, after extracting candidate boats through clustering, we generate models based on boat dimensions and fit them to clusters. Furthermore, we identify static structures and dynamic boats based on position changes over time. The proposed methodology enables stable detection of surrounding boats in urban river environments. It can distinguish between static structures such as bridges and embankments and actual boats in motion. It also allows for the stable recognition of surrounding boats. In the experimental phase, a battery-powered boat equipped with LiDAR navigated along designated routes on the Kanda River and Nihonbashi River. The detection process was performed at a frequency of 1 Hz, and the success rate varied by boat, ranging from 41.18% to 100.00%. Despite these fluctuations, the study demonstrates that model-based moving object recognition using LiDAR and scan matching has the potential to improve autonomous navigation systems. This research contributes to advanced maritime navigation through autonomous control to enhance safety and reduce the human workload.*

Keywords: *SLAM, scan matching, LiDAR, autonomous boats, urban rivers*

Introduction

In recent years, with the rapid development of information and communication and sensing technologies, research and development on autonomous boats has become active worldwide. Large-scale demonstration experiments aimed at the social implementation of unmanned boats are being conducted mainly in Europe and the United States, attracting attention from the perspectives of safety, efficiency, and labor savings. In Japan as well, various technological developments are underway, including automatic route-following control systems, automatic berthing systems utilizing global navigation satellite system (GNSS) and light detection and ranging (LiDAR), and automatic collision avoidance navigation systems capable of operating day and night. However, collision accidents involving small boats such as fishing and pleasure boats remain prevalent. One reason for this is that, unlike large boats such as tankers and container ships, many small boats are not equipped with advanced navigation assistance devices and rely heavily on the visual navigation skills of their operators. Crew members of small boats are forced to perform navigation tasks, including constant surveillance of their surroundings, for extended periods of time, which imposes a heavy burden and contributes to accidents caused by human error. Additionally, in metropolitan areas such as Tokyo, population concentration has led to severe traffic congestion on land, prompting efforts to utilize river transportation as a solution. In fact, the Tokyo Metropolitan Government is promoting policies to use rivers for commuting and school routes, resulting in an increasing trend of boat traffic on urban rivers. However, urban rivers are narrow, with an average width of about 20 meters, and are accompanied by many fixed structures such as bridge piers and embankments. In addition, many small and pleasure boats pass through, so technology that can accurately recognize surrounding boats and obstacles is essential for safe navigation. Conventionally, collision avoidance between boats has relied on position information sharing using GNSS and image processing based on deep learning techniques such as faster region-based convolutional neural networks. However, there is a limitation in that GNSS-based position sharing does not function in non-GNSS positioning environments such as under bridges. Furthermore, although deep learning-based image recognition is highly accurate, it requires large amounts of training data to recognize boats with different directions and shapes, leading to challenges such as insufficient training data and limitations in generalization performance.

Therefore, to overcome these challenges in this study, we propose a new method to detect surrounding boats using a model-based moving object recognition technique based on scan matching, utilizing LiDAR installed on boats. This method enables the recognition of dynamic boats independent of the environment and is effective even in non-GNSS environments. Furthermore, this study conducts experiments on urban rivers using a boat equipped with LiDAR to evaluate the effectiveness of the purposed avoidance method.

Literature Review

Research on detecting and tracking moving objects using LiDAR on mobile platforms is diverse, including the tracking of moving objects using LiDAR mounted on motorcycles. In this study, we propose a method for detecting moving objects using in-vehicle LiDAR, combining self-localization estimation and occupancy grid difference. First, the vehicle's position is sequentially estimated using moving least squares registration (MLSR) based on LiDAR scans. The obtained point clouds are aligned with the map coordinate system. MLSR is a method used for aligning LiDAR point clouds, which estimates the self-position by locally aligning sequentially obtained scans with a known map or the previous frame using least squares. A key feature is its robustness to noise, which is achieved through local smoothing, enabling stable alignment even in a moving sensor environment. This makes it possible to eliminate apparent motion caused by vehicle movement. Next, the aligned point cloud is converted into an occupancy grid map, and moving cells are extracted by comparing the occupancy states of between consecutive frames. After detecting moving objects, a tracking algorithm is applied to cluster and match the objects in a time series. The position and velocity vectors of each object are updated, and continuous trajectories are generated by identifying the same objects between frames. This process enables the continuous tracking of moving objects even in the event of temporary occlusion or observation gaps, thereby allowing allows for stable recognition in traffic environments. The experiment used point clouds obtained from LiDAR mounted on a vehicle to verify the road environment of a university campus. Moving objects such as pedestrians and other vehicles could be clearly extracted using grid difference, and their movements could be reliably tracked through clustering and tracking. In particular, tracking was possible even for vehicles at a certain distance or multiple pedestrians, demonstrating the effectiveness of this method. On the other hand, accuracy degradation in complex traffic environments has been reported as an issue. For example, when multiple moving objects are in proximity, clustering errors are likely to occur, and tracking may be interrupted because of temporary obstruction. Furthermore, there are limitations in estimating the size and shape of moving objects, and false detections have been attributed to sensor noise and ground reflection.

Considering these issues, it would be beneficial to introduce sensor fusion with cameras and radars in addition to LiDAR processing alone, and to integrate it with deep learning-based object recognition methods to improve classification accuracy and robustness. Furthermore, reducing computational costs and ensuring real-time performance are also important issues to be considered for practical application. This study demonstrates the effectiveness of a moving object recognition method using LiDAR mounted on a mobile platform and explores the potential applications of this technology in future autonomous driving and driver assistance systems.

Methodology

The processing flow of the proposed method is shown in Figure 1.

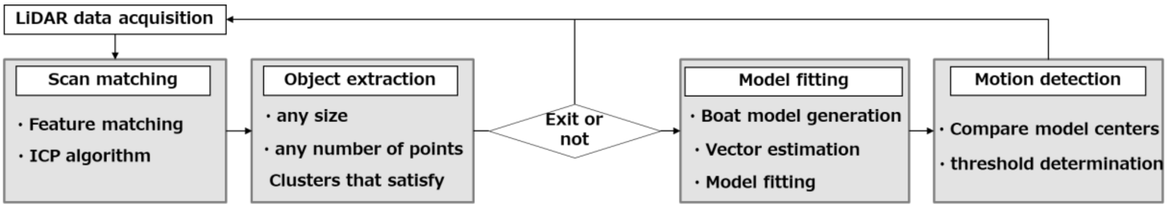

Figure 1: Proposed Methodology.

a. Scan matching:

Figure 1 shows an example of scan matching, in which point clouds are aligned between an arbitrary scan (the current scan) and previous scan (the reference scan). Variation in point intervals within each scan greatly affects the alignment accuracy in scan matching, so the scan point intervals are homogenized as a preprocessing step. For initial posture estimation, a feature-based matching algorithm using fast point feature histograms (FPFH) is applied. FPFH is a statistical method that describes the angular relationships and geometric configurations between the normal vectors of neighboring points at each point. This feature is robust to rotation and translation, and introduces an efficient mechanism for expressing the relative relationships between neighboring points to reduce the computational load. As a result, FPFH quickly and accurately identifies the local shape of point clouds, and is extremely effective in establishing a rough correspondence between scans taken at different times. In particular, in environments where the sensor moves, there is a tendency for the initial positional error to be large. However, FPFH makes global initial pose estimation possible, which stabilizes subsequent alignment processing. After the initial pose estimation, the iterative closest point (ICP) algorithm is applied to achieve higher-precision alignment. While conventional point-to-point ICP minimizes the Euclidean distance between corresponding points, point-to-plane ICP considers the normal direction on the reference scan. It then minimizes the distance between the corresponding points and their respective tangent planes. This improves convergence speed, especially in environments with many planar shapes or smooth surfaces, and also enhances the alignment accuracy. Furthermore, point-to-plane ICP is relatively robust against noise and can operate stably even with incomplete point clouds obtained in real environments. As mentioned earlier, this study achieves fast and accurate scan matching by combining a global initial pose estimation using FPFH and local refinement using point-to-plane ICP.

b. Object extraction:

As shown in Figure 1, clustering is first performed on both the current and reference scan point clouds after scan matching. In a waterborne mobile mapping system (MMS). point clouds of the water surface caused by waves around a boat can be measured, as shown in Figure 2. Because these water surface point clouds adversely affect boat recognition, the point clouds at the bottom of each cluster are removed as a preprocessing step.

Figure 2: Point Clouds of Boats and Water Surfaces (Side View).

Next, only those clusters whose bounding box size and number of points satisfy predetermined conditions are extracted from the current scan. This process rejects noise clusters with an extremely small number of points and targets only clusters that are of a reasonable size for boats. As a result, unnecessary processing is avoided, and the computational load for boat detection is reduced. After that, the extracted clusters are matched with each other. The closest cluster in the reference scan is assigned to each cluster center point in the current scan as the corresponding cluster. However, if the distance between the center points of the corresponding clusters is smaller than the threshold, or if the difference in the number of points exceeds the threshold, the cluster is excluded from further processing and removed from the focus. In scan matching, only static objects have two point clouds that match perfectly. point sets corresponding to dynamic objects do not match. In addition, the difference in the number of points between corresponding clusters can become extremely large. This occurs when clustering fails as a result of mismatches in the point set structure between time series for structures with complex shapes, such as seawalls and bridges. Therefore, clusters with extremely small center point distances or extremely large differences in the number of points can be regarded as static objects and excluded from the detection target.

c. Model fitting:

The model fitting process starts with the generation of a boat model. Next, the overall length, beam, and height above the waterline are calculated based on the size of the extracted cluster. Then, a hull point cloud model suitable for shape model fitting is created. An example of boat model generation is shown in Figure 3.

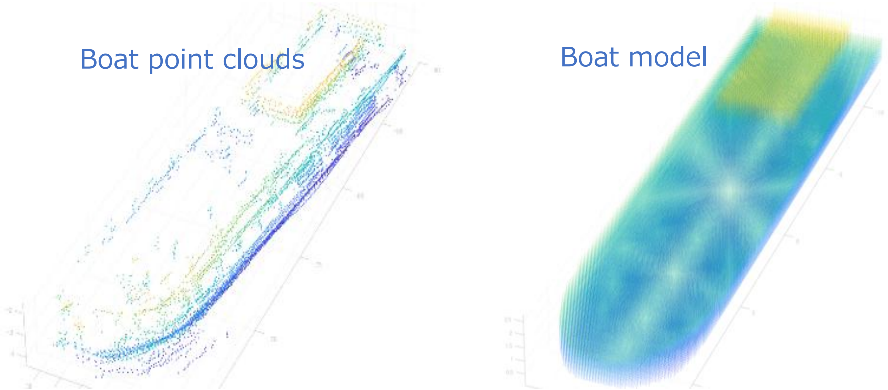

Figure 3: Example of Boat Model Generation (Left Figure: Actual Boat Point Clouds, Right Figure: Boat Model).

The main purpose of introducing model fitting into this study is to estimate the center position of clusters more stably and accurately. Although averaging the coordinates of the point clouds within a cluster or setting a bounding box to determine its geometric center are possible methodologies, they are not suitable for LiDAR measurements of urban rivers. These measurements often include complex structures such as bridge piers and riverbanks, as well as unwanted point clouds caused by reflections from the water surface, and partial obstructions. These issues often result in incomplete and biased distributions of cluster point clouds. Under such conditions, performing simple center estimation can misidentify a position significantly offset from the actual center of gravity of the boat as the center. Especially, estimation based on averages or bounding boxes becomes highly unstable, when only part of the boat's hull is measured or when point clouds from surrounding structures are mistakenly included in the cluster. This can result in the detection of incorrect boat positions. These errors can lead to critical errors in object tracking or collision avoidance decisions. By contrast, this methodology generates a model that reflects the boat's shape and optimally fits it to the cluster in advance. This process minimizes the impact of point cloud defects and disturbances, thereby enabling more stable estimation of the boat's center position and attitude. In other words, the model fitting enables estimation of the shape of the entire hull, even if the cluster point clouds are incomplete, thereby achieving robust boat recognition.

Next, the generated boat model is fit to the cluster. Figure 4 shows an example of boat model fitting. The rotation matrix and translation vector are calculated using the model based on the bow direction vector and center coordinates of the target cluster. Then, the model is aligned. The bow direction vector of the boat model is calculated using the known bow and stern coordinates. On the other hand, the bow direction vector of the target cluster is calculated using the center coordinates of the current cluster and the corresponding cluster from the previous scene. In other words, the boat's direction of travel is estimated by comparing the positions of the cluster center at a given time to their positions at the previous time. This estimated direction is used as the bow direction vector. Incorporating model fitting enables robust estimation of the boat's center even in the complex environment of urban rivers, which improves the reliability of subsequent moving object recognition and collision avoidance.

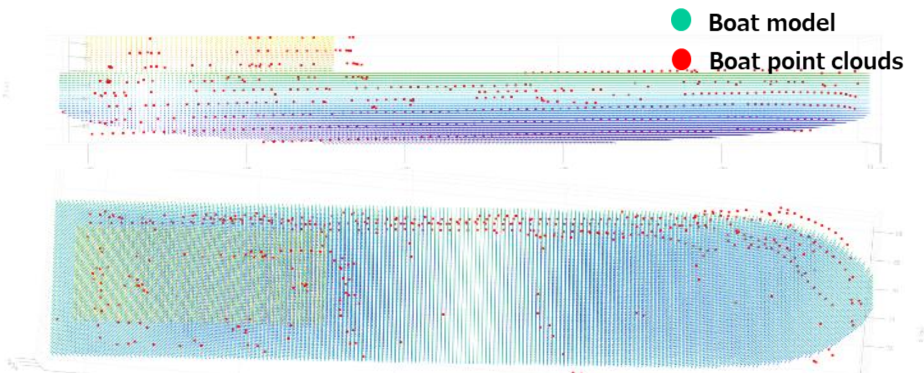

Figure 4: Examples of Fitting Boat Models (Top Figure: Side View, Bottom Figure: Bird's-Eye View).

d. Boat detection:

As shown in Figure 1, if the center-to-center distance of the corresponding boat model is within the set threshold, the boat is judged to be a "static object." If the distance exceeds the threshold, the boat is recognized as a "dynamic object."

Experiments

We conducted experiments using LiDAR (VLP32C, Velodyne) and an omnidirectional camera (Ladybug 5+, FLIR) mounted on a battery-powered boat (Raicho I) as an MMS to acquire point clouds of landmarks and other boats (Figure 5).

Table 1: Specifications of the LiDAR(VLP-32C).

Size	Diameter 103mm, height 86.9mm
Weight	925g
The Number of channels	32
Measurement range	200m
Distance measurement accuracy	Up to±3cm (Typical)
Vertical field of view	40°(-25°to ±15°)
Vertical angle resolution	0.33°
Horizontal field of view	360°
Horizontal angle resolution	0.1°-0.4°
Scanning speed	5Hz-20Hz

Table 2: Specifications of the Ladybug 5+.

Size	Diameter 197mm, height 160mm
Weight	3.0kg
Resolution	2048×2464
Frame rate	30
Sensor	Sony IMX264, CMOS, 2/3”
Shutter	Global shutter
Dynamic range	70dB
Resolution	2 mm@10 m

Figure 5: Overall View of the Experimental Measurement System.

a. Platform

In this study, to conduct experiments in narrow rivers safely, the utilized battery-powered boat was small length 10 meters). Moreover, battery-powered propulsion was considered suitable for stable point cloud acquisition because of the associated low rocking and vibration levels.

Table 3: Specifications of "Raicho I".

Length	Approx. 10m
Width	Approx. 2.3m
Height of underwater	Approx. 1.2m
Motor power	25kW
Gross ton	2.5ton
Maximum speed	10kn
Boat type	Battery-powered propulsion, quick-charge capability
The number of passengers	Crews: 2, Passengers: 10

The experiment was conducted along the Kanda River and Nihonbashi River. The route is shown in Figure 6. The Kanda River consists of an open sky environment. Therefore, LiDAR was used to acquire point clouds of buildings and revetments for stable scan matching along the river with GNSS data. Part of the Nihonbashi River is covered by a highway. The river under the elevated highway has a complex shape because of numerous ramps and interchanges. Although the Nihonbashi River is wide, it has sections suitable for measuring oncoming boats owing to numerous road bridges. We observed six other boats (Boats 1-6) in the experimental sections, as shown in Figure 6. Boat 1 was observed near the Suimon Terrace Liaison Boat 2 near the Eitai Bridge, Boat 3 near the Sumida River Bridge, Boat 4 near Kiyosubashi Bridge, Boat 5 near Ryogoku Bridge, and Boat 6 near Manaita Bridge.

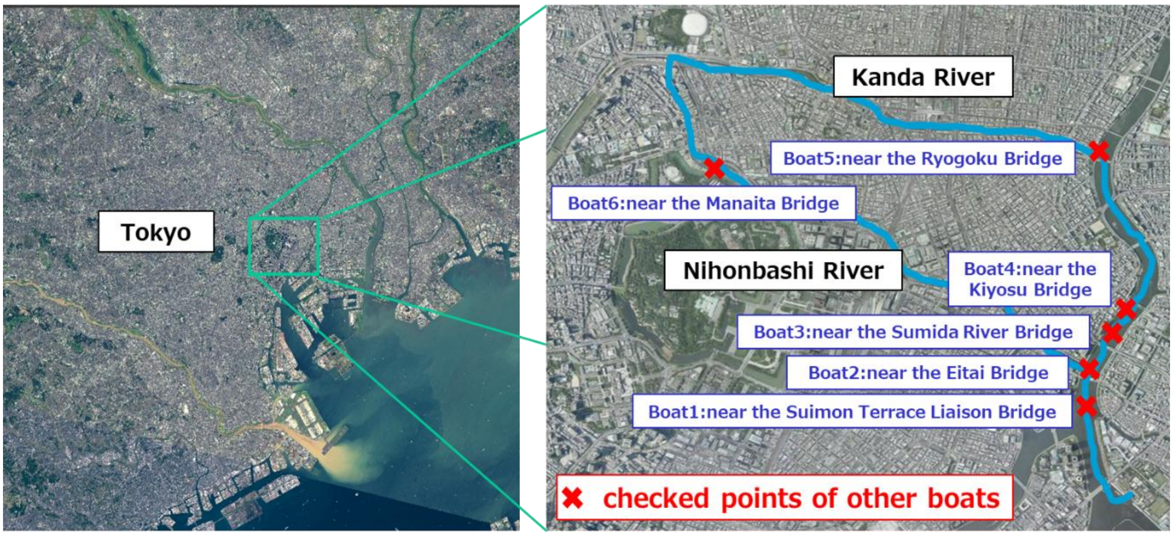

Figure 6: Location of the Experimental Site.

Results and Discussion

Figure 7 shows the measurement scene from the boat in our experiment. Figure 8 shows an example of successful boat detection. During the detection process, boats were displayed in red and the background was displayed in blue.

Figure 7: Boat Measurement Scene.

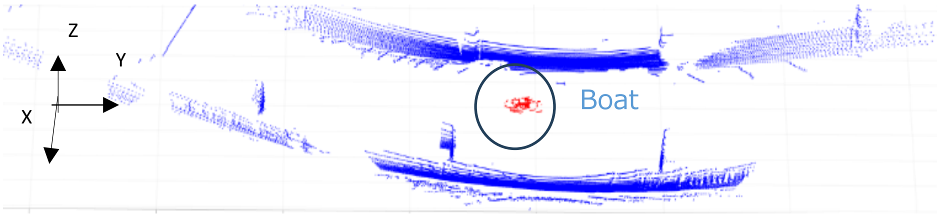

Figure 8: Successful Examples of Boat Detection.

The boat detection process was performed every 10 frames at a frequency of 1 Hz. When the time series difference between frames was small, the shift in the boat cluster after scan matching also become small, which affect the detection accuracy. Additionally, because boat detection processing for one frame took approximately 1 second, real-time processing could not be maintained. Therefore, to ensure sufficient processing time, the time series difference between frames undergoing processing was increased. Variation was observed for each boat. The detection rate for six boats, which was calculated using Equation 1, is shown in Table1. The highest detection rate was 100.00%, and the lowest was 41.18%.

$$Detection\ rate = \frac{The\ number\ of\ frames\ in\ which\ boats\ were\ correctly\ detected}{The\ number\ of\ frames\ showing\ boats\ underway}$$

(Equation 1)

Table 4: Boat Detection Rate.

	Boat1	Boat2	Boat3	Boat4	Boat5	Boat6	Average
Detection Rate[%]	64.00	66,67	100.0	41.18	80.95	71.05	70.64
Total Frames	25	27	31	34	42	38	33

A successful example of scan matching is shown in Figure 9. While static structures such as bridges and embankments matched the boats were misaligned.

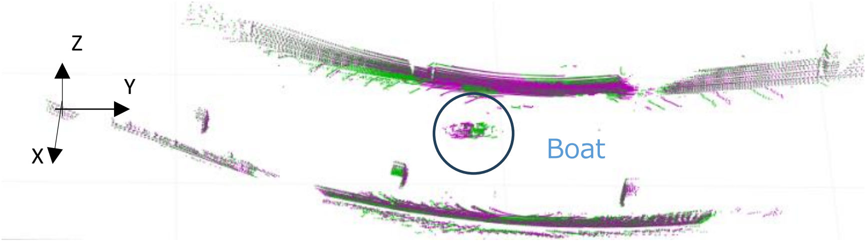

Figure 9: Successful Examples of Scan Matching.

The scan matching processing was performed on frames where boats were present. Table 2 shows the success rate of scan matching for scenes with six boats. The success rate was calculated using Equation 2, and success was determined visually. Also scan matching was stable in many frames, and a significant decrease in the success rate was observed for Boat 1.

$$success\ rate = \frac{Number\ of\ frames\ that\ matched\ correctly}{The\ number\ of\ frames\ showing\ boats\ underway}$$

(Equation 2)

Table 5: Scan Matching Success Rate.

	Boat1	Boat2	Boat3	Boat4	Boat5	Boat6
Success Rate[%]	84.00	100.0	100.0	100.0	100.0	100.0
Total Frames	25	27	31	34	42	38

The average detection rate for the boats was 70.64%, and undetected boats were observed in all cases. Possible causes for this include variations in the accuracy of model fitting and scan matching. Figure 10 shows an example of model fitting failure, which was the result of an error that occurred in the estimation of the bow direction vector of the target cluster, as shown in Figure 11. The bow direction vector is calculated using the center coordinates of the target cluster and the corresponding cluster from the scan 1 second earlier. The coordinates used here are relative positions based on the boat's coordinate system in each scan, as opposed to absolute coordinates on the map. Therefore, because this study assumed that the boat was moving, it was not possible to estimate the movement of the target cluster between scans accurately. To achieved this, it is essential to estimate the position sequentially for each LiDAR scan and align the acquired point cloud with the map coordinate system. This allows the position of the target cluster to be expressed in consistent absolute coordinates, thereby improving the accuracy of the direction of travel estimation.

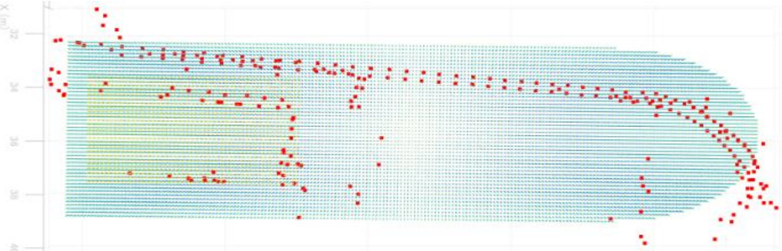

Figure 10: Examples of Model Fitting Failures.

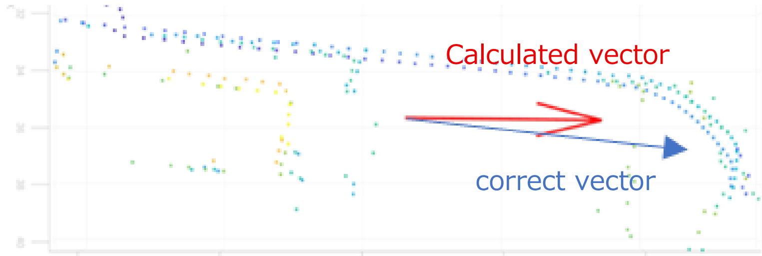

Figure 11: Incorrect Estimation of Bow Direction Vector.

Figure 12 shows an example of scan matching failure observed in the Boat 1 scene. In this example, the failure occurred as a result of the extremely small number of landmark point clouds. In aquatic environments such as urban rivers and oceans, feature-poor surroundings are inevitable. Therefore, it is necessary to supplement the estimation of position and attitude changes by fusing data from an inertial measurement unit (IMU), which can accurately measure boat's short-term acceleration and angular velocity, thereby enabling prior estimation of relative attitude and position changes between scans. Providing these estimated values as initial values for scan matching is expected to improve convergence stability and accuracy even in cases involving few landmarks and unstable point cloud alignment.

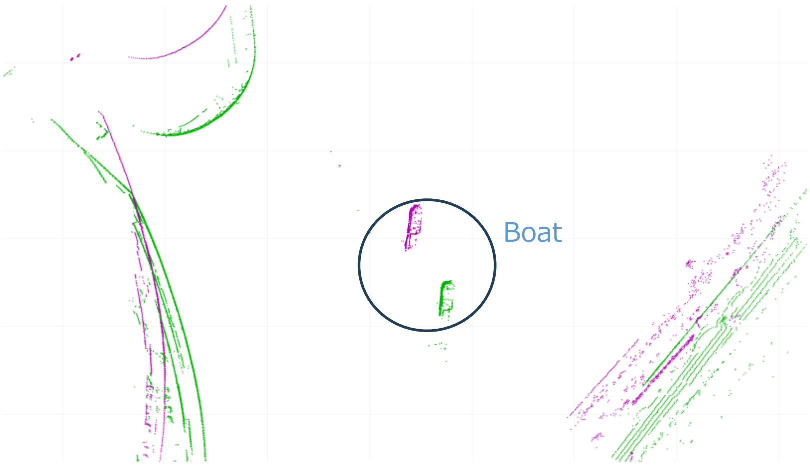

Figure 12: Examples of Scan Matching Failures.

Conclusion

This study proposed a model-based boat recognition methodology for urban river navigation using LiDAR and scan matching. This approach addresses the limitations of conventional methods, such as GNSS-based position sharing and image recognition based on deep learning, which face difficulties in non-GNSS environments and require extensive training data to handle the wide diversity of boat shapes and orientations. By combining FPFH for global initial alignment and point-to-plane ICP for local refinement, the proposed method enables stable and accurate scan matching. Subsequently, clustering and model fitting were applied to extract boat candidates, generate hull models, and estimate boat centers more reliably, even under complex riverine conditions that include reflections, partial occlusions, and structural interference. Experiments were conducted using a battery-powered boat equipped with LiDAR and a camera on the Kanda River and Nihonbashi River, which include diverse environments such as open-sky and bridge-covered sections. The detection process, executed at 1 Hz, achieved a boat recognition rate ranging from 41.18% to 100.00% depending on the observed vessel, with an average detection rate of 70.64%. The success rate of scan matching was generally high, although occasional failures occurred in environments with limited landmarks. These results demonstrated the effectiveness of the proposed methodology in distinguishing between static structures and moving boats, thereby contributing to the reliable recognition of dynamic objects for safe navigation. However, some challenges remain, particularly in regard to failures in model fitting due to the inaccurate estimation of bow direction vectors and scan matching instability in feature-poor aquatic environments. In future work IMU-based motion estimation should be integrated to enhance robustness, and evaluations should be expanded by increasing the number and variety of boat samples to validate applicability further. Overall, the present findings indicate the potential of model-based moving object recognition with LiDAR and scan matching to support advanced autonomous navigation, improve safety in urban rivers, and reduce the workload on human operators.

References

Shinbo, M., Hirasawa, M., and Oshima, M., (2005). Extraction of Navigational Boats from Time-Series Images of Navigational Environment Acquired by Multi-sensors. Transactions of the Nautical Institute of Japan, No.114, pp.65-72.

Zhou, L., Shinbo, M., and Oshima, M., (2007). Real-Time Boat Extraction Method from Time-Series Images of Navigational Environment. Transactions of the Nautical Institute of Japan, No.117, pp.175-182.

Matsumoto, Y., (2013). Ship Image Recognition Using HOG Features. Transactions of the Nautical Institute of Japan, No.129, pp.105-112.

Kondo, M., Miyashita, T., Imai, R., Inaishi, M., Furuya, M., and Nakagawa, M., (2020). A Study on Ship Tracking and Course Prediction Using Panorama Camera -For River Traffic-. Transactions of the Japan Society of Navigation Engineers, No.143, pp.31-41.

Muro, S., Hashimoto, M., Takahashi, K., (2020). Moving-Object Tracking Using Lidar Mounted on Motorcycle in GNSS-Denied Environments. Transactions of the Harris Science Review of Doshisha University, No. 60, pp. 214-221.