Title	A Real-Time Semantic-Aware Simultaneous Localization and Mapping for Unmanned Aerial Vehicles
Author(s)	Nguyen Canh, Thanh
Citation	
Issue Date	2024-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19354
Rights	
Description	Supervisor: 丁 洛榮, 先端科学技術研究科, 修士(情報科学)

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

A Real-Time Semantic-Aware Simultaneous Localization and Mapping for
Unmanned Aerial Vehicles

2210429 Nguyen Canh Thanh

Unmanned Aerial Vehicles (UAVs) are essential in various fields such as
disaster monitoring, environmental surveys, search and rescue missions, and
infrastructure inspection, where a precise understanding of the surrounding
environments is crucial. A fundamental challenge for UAVs operating in dy-
namic and unstructured environments is the capability to navigate and map
their surroundings in real-time accurately. Simultaneous Localization and
Mapping (SLAM) is an essential technology that addresses this challenge by
enabling UAVs to construct detailed maps while simultaneously determining
their location within these maps. Visual SLAM has become increasingly cru-
cial in the effective localization and representation of a map consisting of 3D
points; however, it lacks semantic information and serves for high-level tasks.
Numerous previous approaches have aimed to build dense maps; however,
these reconstructions are still just aggregates of points, and thus lack any
supplementary semantic information or relationships. A robot, moreover,
must be capable of mapping its environment, localizing itself within that
map, and comprehending the semantic information of the surrounding scene.
Semantic SLAM addresses three fundamental tasks simultaneously, aiming
to produce the most precise and comprehensive environmental model in an
environment. Achieving this requires a careful balance between the accuracy
of the semantic map and the memory resources it consumes. Over recent
decades, semantic SLAM has garnered increasing interest and has been ex-
plored in various ways by different research communities, driven by the goal
of practical deployment in a real-world application. The broad interest has
expanded the problem's scope and provided diverse perspectives, leading to
numerous approaches based on various theories and concepts. However, this
has also created a disconnect between research paths that could be mutually
beneficial. However, integrating localization, semantic segmentation, and 3D
reconstruction simultaneously poses a significant challenge, particularly for
UAVs that operate with limited power and computational capacities.

To mitigate these challenges, this thesis emphasizes the development of
Semantic SLAM systems that integrate metric environment structures with
semantic object information to create comprehensive semantic maps. Our
proposed method is organized into two key enhancements: a 3D semantic
mapping method and a 2.5D probabilistic metric map approach. By inte-
grating semantic details, we aim to enhance the effectiveness of the semantic
SLAM system. Initially, we introduce an innovative strategy to tackle the

issues related to the extraction and use of semantic data in UAV operations. Our framework combines cutting-edge visual SLAM for accurate 6-DoF pose estimation with sophisticated object segmentation techniques at the back end. To enhance the framework's computational and storage efficiency, we employ a simplified voxel-based 3D map representation known as OctoMap for system construction. Additionally, we integrate a fusion algorithm to retrieve semantic information from each frame in the front-end SLAM task and the associated point. Secondly, we propose to construct a probabilistic metric map enriched with object data from RGB-D images. This method integrates a cutting-edge YOLOv8-based object detection framework upfront and a 2D SLAM method - Cartographer, at the back end. To track and position semantic object categories obtained from the front-end interface, we utilize the innovative BoT-SORT methodology. A new association technique is proposed to determine objects' positions and project them onto the metric map. Unlike previous studies, our method focuses on navigating environments that contain various hollow objects on the bottom. The output is a probabilistic map that significantly enhances the representation of the map by incorporating object-specific details, including class distinctions, accurate positions, and object heights.

To demonstrate the pose estimation performance of our semantic SLAM system, we performed evaluations using two different types of datasets: 1) publicly available TUM real-world RGB-D data sequences and 2) a Gazebo simulation dataset. We assessed the precision of 6-DoF pose estimation using the Root Mean Square Error (RMSE) of Absolute Trajectory Error (ATE) and Relative Pose Error (RPE). Compared to the state-of-the-art visual SLAM - ORB-SLAM2, the results demonstrated precise pose estimation and smooth movement within environments. Our mapping outcomes show that our system greatly enhances mapping accuracy, computational efficiency, and the UAVs' capability to autonomously navigate complex scenarios. Additionally, our system is tested on an embedded computer - Jetson Xavier AGX unit, to illustrate its effectiveness in real-world applications.

In summary, this thesis presents the development of advanced Semantic SLAM systems designed to enhance the autonomous capabilities of UAVs in dynamic and unstructured environments. By integrating semantic understanding and probabilistic mapping methodologies, our approach significantly improves mapping accuracy, computational efficiency, and UAV navigation. Future work could focus on building a sustainable 3D active semantic SLAM system that provides essential information for UAV applications, ensuring the safe operation of flying equipment, and enhancing localization capabilities. This approach takes advantage of the full potential of camera data, making additional sensors such as GPS or LiDAR unnecessary.