CHAPTER 1
INTRODUCTION

This chapter provides a brief overview of the research. It consists of six sections, starting with background, problem identification, objectives, scope of work, research methodology, and the structure of this thesis. A more detailed explanation will be provided in the subsequent chapters.

1.1 Background

Semantic mapping combines geometric and object-level information to give robots a more comprehensive understanding of their environment. Unlike traditional SLAM methods, which only provide geometric maps, semantic mapping allows robots to identify and label objects, enabling more intelligent decision-making [5, 6]. Object detection, especially through deep learning models, has made this possible by identifying and classifying objects in real-time [4]. By integrating object detection with SLAM, robots can navigate environments with greater awareness, identifying important objects such as doors, obstacles, or tools [7]. This is especially important in scenarios where the robot must explore new environments and identify explorable areas for efficient navigation and task completion [6–8]. The fusion of camera, odometry, and Light Detection and Ranging (LiDAR) data has proven particularly useful, as cameras provide rich visual information, while LiDAR delivers precise distance measurements and odometry provides accurate localization. .

While LiDAR excels at providing accurate depth measurements and mapping the structural geometry of an environment, it suffers from limitations such as difficulty in detecting object textures and recognizing object types. LiDAR systems often struggle with materials that absorb or reflect light poorly, such as glass or shiny surfaces, which can result in incomplete or inaccurate data. Moreover, LiDAR's inability to capture color and texture information reduces its capability for detailed object recognition [9].

On the other hand, cameras are excellent at capturing detailed visual information, such as color and texture, which are essential for object recognition and classification. However, cameras are affected by environmental factors such as lighting conditions and occlusions, making it difficult for them to provide accurate distance measurements or operate effectively in low-light conditions. This limitation reduces their effectiveness in tasks that require precise localization, such as navigation and mapping [9, 10].

The complementary nature of LiDAR, cameras, and odometry has driven the development of sensor fusion techniques, which combine the strengths of these sensors to provide richer and more accurate environmental data. By fusing LiDAR's depth perception, the camera's visual recognition capabilities, and odometry's localization accuracy, robots can generate a more holistic understanding of their surroundings, allowing for both geometric mapping and detailed object labeling [11]. This integrated approach enriches the environment's representation with both spatial and semantic information, addressing the limitations of each sensor alone, and enabling more accurate identification of explorable areas [12].

Recent works have explored the integration of object detection, odometry, and SLAM for semantic mapping. For example, one study developed a framework that combined a convolutional neural network (CNN) for object detection with 2D LiDAR for mapping [13]. However, the system struggled to achieve real-time performance due to the computational load of the object detection model, making it unsuitable for time-sensitive applications. Similarly, another study integrated object recognition with SLAM but faced difficulties in handling dynamic environments, leading to inconsistent object placement on the map and inaccurate identification of explorable regions [14]. These studies highlighted the potential of combining SLAM, odometry, and object detection, but neither fully addressed the challenge of achieving real-time performance in complex, dynamic settings.

This thesis seeks to develop an optimized semantic mapping framework that combines object detection and SLAM, specifically tailored for real-time operation using odometry for accurate localization and sensor fusion. The goal is to enhance robotic navigation and environmental understanding by enabling real-time object detection and mapping of explorable areas, particularly in dynamic indoor settings. Through this approach, the research demonstrates that efficient and accurate real-time semantic mapping is possible without sacrificing computational efficiency.

1.2 Problem Identification

The primary issue identified in this research involves the limitations of integrating LiDAR, odometry, and camera data for semantic mapping, especially in real-time applications. While LiDAR provides accurate distance measurements and geometric information, it lacks the rich visual and contextual information offered by cameras, making it difficult for systems to fully understand the environment and determine which areas are explorable. On the other hand, cameras excel in visual object detection but struggle in low-light conditions or when precise distance measurements are required. Odometry provides accurate localization but can accumulate errors over time, especially in complex environments. Another challenge is the high computational load that occurs when combining object detection with SLAM. These limitations highlight the need for optimized algorithms that can efficiently manage the fusion of sensor data and SLAM while maintaining high accuracy in identifying explorable areas and ensuring real-time performance.

1.3 Objectives

To simplify the explanation, this thesis assumes the following objectives:

1. Develop a real-time semantic mapping system integrating object detection and SLAM using camera and LiDAR data, with a focus on identifying explorable areas.
2. Focus on efficient fusion of camera and LiDAR data to enhance both geometric and semantic mapping, particularly in detecting areas that are navigable or require exploration.
3. Utilize deep learning models like YOLOv3 for accurate real-time object detection within the SLAM framework.
4. Ensure the system performs effectively in indoor environments with real-time processing, enabling the robot to navigate through and map explorable areas.

1.4 Scope of Work

In this research, the defined limitations are presented as follows:

1. The thesis focuses on developing a real-time semantic mapping framework that integrates object detection and SLAM using both camera and LiDAR data to identify and map explorable areas.
2. The research will address the fusion of visual and distance data from the camera and LiDAR to enhance mapping accuracy, context, and the identification of navigable regions.

3. The object detection component will be implemented using the YOLOv3 model.
4. The scope is limited to indoor environments, where the system will be tested under various conditions to validate its ability to map explorable areas in real-time.
5. The thesis will not explore outdoor environments or focus on 3D SLAM methods, with the primary emphasis on 2D mapping and object detection in indoor spaces.

1.5 Research Methodology

This thesis is divided into work packages (WP).

- **WP 1:** Proposed System and System Design. This work package covers the design and architecture of the proposed real-time semantic mapping framework. The system is designed to integrate object detection and SLAM, utilizing both camera and LiDAR data to generate a more comprehensive map of the environment, with a specific focus on mapping explorable areas. The focus will be on developing a modular, scalable system architecture that can efficiently handle real-time processing.
- **WP 2:** Implementation and Integration of Algorithms. In this stage, the individual algorithms for object detection (YOLOv3) and SLAM (Gmapping) will be developed and implemented. Special attention will be given to the integration of the camera and LiDAR data, ensuring that sensor fusion enhances both geometric and semantic understanding of the environment, particularly in identifying explorable areas.
- **WP 3:** Experimental Setup and Testing. This work package will define the experimental procedures for validating the system. It includes setting up the hardware environment, including the camera, and LiDAR. The system will be tested under various indoor conditions to evaluate its real-time performance, accuracy, and robustness, particularly in mapping explorable areas. This section will also detail the metrics and methods used to assess the performance of the system in different scenarios.
- **WP 4:** Semantic Mapping Construction: This work package focuses on the development of the semantic mapping process, which combines object detection with SLAM to create a map enriched with semantic labels. The aim

is to provide both geometric and object-level context information about the surroundings and identify explorable areas. This WP will explore methods for integrating object detection into the map, ensuring that the system can dynamically update the map with object labels and navigable regions as the environment changes.

- **WP 5:** Experimental Results and Analysis. The final work package will analyze the experimental results obtained from the previous testing stage. This section will evaluate the system's overall performance based on metrics such as accuracy of object detection, mapping precision, real-time processing capabilities, and computational efficiency. Special attention will be given to how well the system can map explorable areas. The results will be compared to existing methods to highlight improvements and potential areas for further optimization.

1.6 Structure of this Thesis

The structure of this thesis is organized as follows:

- **Chapter 1 - Introduction:** This chapter provides an overview of the research, including the background, problem identification, objectives, scope of work, research methodology, and the structure of the thesis itself.

- **Chapter 2 - Basic Concept:** This chapter discusses key theoretical concepts related to the research. Topics covered include semantic mapping, SLAM, the Robot Operating System (ROS), object detection, and sensor fusion techniques with camera and LiDAR integration. It also delves into the concept of explorable areas, discussing its importance in the context of robotic navigation and mapping.

- **Chapter 3 - Methodology and Technique:** This chapter outlines the system design, explaining the integration of sensors and algorithms used for object detection, SLAM, and semantic mapping, with a focus on identifying explorable areas. Detailed descriptions of hardware and software configurations, including the use of ROS, YOLOv3, and Gmapping, are provided.

- **Chapter 4 - Experiment Results:** This chapter presents the results of experiments, which include map construction using SLAM, real-time object detection using YOLOv3, and the integration of these components into a semantic map with emphasis on explorable areas. The performance of the system is evaluated based on metrics such as accuracy, real-time processing, object detection precision, and the ability to map navigable areas.

- **Chapter 5 - Conclusion and Future Works:** This final chapter summarizes the key findings of the research and outlines potential areas for future development. Suggestions are made for further improvements, including enhancements to the detection of explorable areas and the possible integration of 3D mapping using depth cameras.

The chapters follow a logical progression, from a theoretical foundation to experimental application and evaluation, culminating in conclusions and proposals for future work. This structure ensures that the research objectives are clearly addressed and the proposed solution is thoroughly examined, with a particular focus on the identification of explorable areas for efficient robotic navigation.