Semantic SLAM and 3D Mapping in Robotics: A Survey

www.surveyx.cn

Abstract

This survey explores the interdisciplinary field of Semantic SLAM, 3D Mapping, and their roles in advancing robotic autonomy. By integrating semantic information into SLAM systems, robots achieve enhanced cognitive and operational capabilities, enabling precise navigation and interaction in complex environments. The survey highlights the pivotal role of advanced algorithms and multi-sensor data fusion in improving the robustness and efficiency of SLAM systems, particularly in dynamic settings. Recent experiments demonstrate significant reductions in drift and improvements in semantic mapping accuracy, underscoring the impact of these integrations. The interdisciplinary nature of this field, encompassing computer vision, robotics, and AI, continues to drive innovations that enhance the functionality and adaptability of robotic systems. Advancements in AI and machine learning, especially deep learning techniques, are poised to further augment the perceptual and navigational capabilities of autonomous systems. The survey emphasizes the critical importance of Semantic SLAM and 3D Mapping in developing intelligent robotic systems capable of autonomous navigation and complex task execution. As the field evolves, the potential for future innovations is vast, promising to unlock new applications and capabilities. The integration of foundation models and exploration of new sensor modalities and algorithmic strategies will be crucial in shaping the future of robotic perception and interaction, paving the way for groundbreaking advancements in this interdisciplinary domain.

1 Introduction

1.1 Interdisciplinary Field Overview

The interdisciplinary nature of Semantic SLAM encompasses computer vision, robotics, artificial intelligence, and cognitive science, which are essential for enhancing mobile robots by integrating semantic information into dense maps, thus overcoming the limitations of geometric and appearance-based data [1]. The combination of 3D semantic reconstruction techniques with geometric features is vital for autonomous navigation, addressing the shortcomings of existing methods [2].

Dense semantic SLAM remains a significant challenge, particularly in autonomous driving applications, emphasizing the need for interdisciplinary collaboration [3]. The innovative use of event cameras—bio-inspired vision sensors that capture pixel-wise intensity changes asynchronously—illustrates the integration of high-speed and high dynamic range technologies in visual odometry and SLAM [4]. Moreover, the roles of visual odometry and SLAM in autonomous navigation and augmented reality further highlight the interdisciplinary essence of Semantic SLAM [5].

In heterogeneous robot teams, real-time decentralized metric-semantic SLAM enables autonomous exploration of 3D environments, showcasing the practical applications of interdisciplinary research in robotics [6]. The application of multimodal approaches, which adapt on-land robotics methods

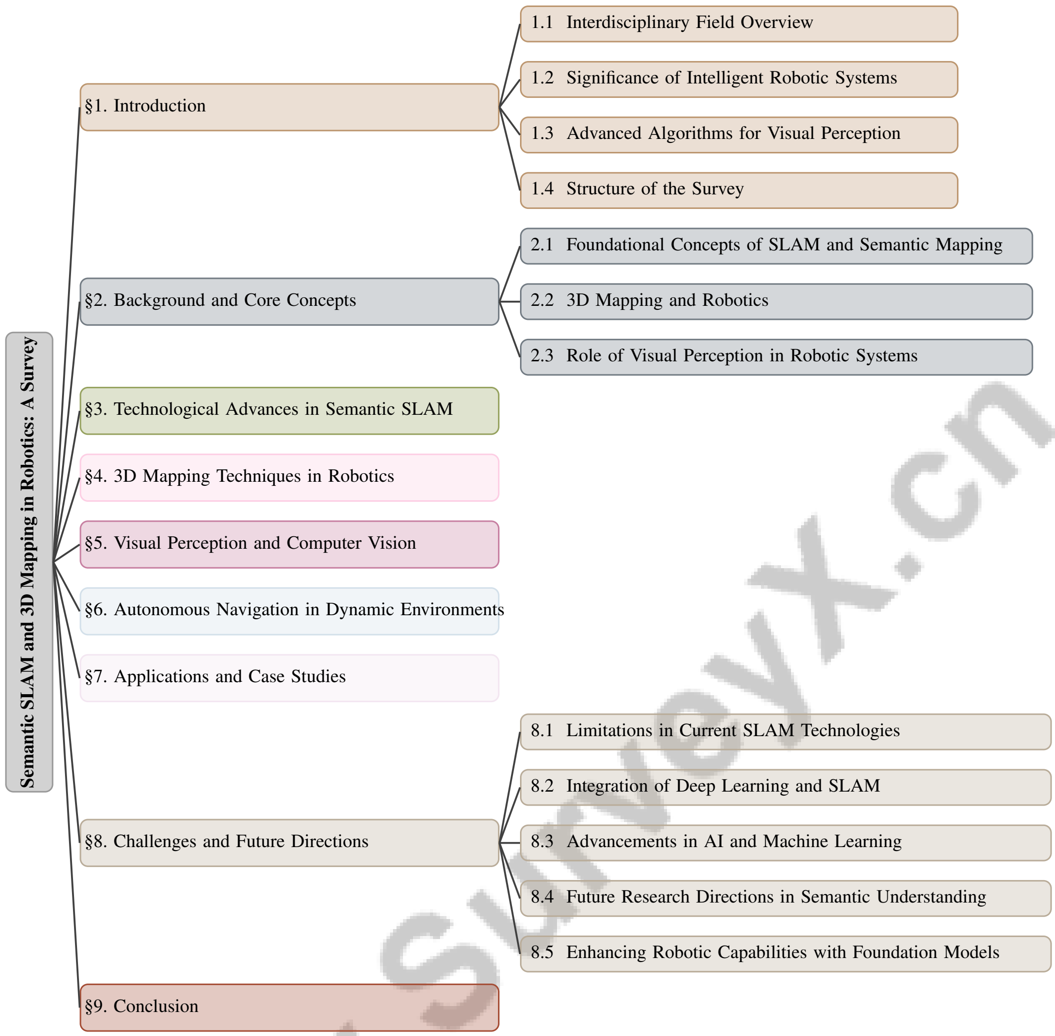


Figure 1: chapter structure

to improve underwater localization, further underscores the interdisciplinary nature of robotics and visual perception [7].

Categorizing existing datasets by their characteristics and applications aids in understanding their suitability for specific perception tasks, thus facilitating advancements in Semantic SLAM [8]. Additionally, addressing semantic understanding in indoor environments is crucial for service robotics, highlighting the importance of integrating semantic segmentation techniques [9].

The rapidly advancing field of Semantic SLAM is driven by diverse technological and scientific innovations that enhance robotic systems' functionality and adaptability while addressing challenges such as object recognition in dynamic environments, data association ambiguities, and modular SLAM architectures. Recent research has introduced robust multi-modal frameworks that improve object feature representation and segmentation accuracy, enabling precise identification of dynamic objects under motion blur. Hierarchical strategies for object association and pose refinement have been proposed to enhance localization accuracy in complex settings, while new architectures promote greater efficiency and code reuse in SLAM implementations [10, 11, 12, 13].

1.2 Significance of Intelligent Robotic Systems

Intelligent robotic systems are crucial in modern technology, enabling navigation and operation in complex, unstructured environments. Their significance is underscored by benchmarks like KITTI-360, which provides comprehensive datasets for comparing models across computer vision,

AI-generated, for reference only.

graphics, and robotics, thereby facilitating research in autonomous driving [14]. The integration of photorealistic mapping within SLAM frameworks, such as PhotoSLAM, exemplifies the importance of these systems in achieving efficient and accurate localization and mapping [15].

Incorporating semantic information into intelligent robotic systems enhances their capability to execute complex tasks outdoors, addressing the limitations of traditional systems reliant solely on geometric data [16]. Scalable and memory-efficient localization approaches, particularly hierarchical visual localization methods, are essential for improving autonomous robot operations [17]. Additionally, advancements in Kimera significantly enhance robustness and accuracy, showcasing superior performance over other state-of-the-art visual-inertial SLAM systems [18].

In healthcare, the integration of intelligent real-time processing of intraoperative Optical Coherence Tomography (iOCT) volumes enhances precision in robotic retinal surgery, highlighting the transformative potential of these systems in medical technology [19]. Although the development of general-purpose robots has been hindered by existing systems' brittleness in unseen environments, ongoing research aims to enhance adaptability and reliability [20].

The exploration and mapping of unknown large-scale environments by UAVs, which ensures obstacle avoidance and accurate pose estimation, illustrate intelligent robotic systems' applications in efficient exploration tasks [21]. Recent methods introducing hierarchical metric-semantic representations combine high-level sparse semantic maps with low-level voxel maps, improving place recognition and loop closure capabilities, thereby demonstrating practical applications in complex environments [6].

Intelligent robotic systems drive innovations in autonomous navigation, optimize operational efficiency, and expand applications across diverse fields. Recent advancements include systems like OK-Robot, which integrates vision-language models for object detection and navigation primitives for movement, achieving a state-of-the-art success rate in open-ended pick-and-drop tasks. Novel planning methods for service robots enable efficient tidying of home environments through learned object placement strategies using multimodal sensor data. Furthermore, the emerging edge computing paradigm enhances multi-robot SLAM capabilities by significantly reducing processing latency and improving real-time map construction, exemplifying the transformative impact of intelligent robotic systems across sectors, from home automation to advanced collaborative robotics [22, 20, 23]. Ongoing research continues to push the boundaries of what these systems can achieve, paving the way for future innovations.

1.3 Advanced Algorithms for Visual Perception

Advanced algorithms for visual perception are crucial in enhancing robotic systems' capabilities to interpret and interact with their environments. The integration of deep learning techniques for semantic segmentation has significantly advanced the field, providing robust algorithms, architectures, and loss functions tailored for various datasets and benchmarks [24]. These techniques enable robots to discern and categorize elements within their environments, facilitating informed decision-making processes.

Spatial attention networks combined with edge detection improve feature selection for visual localization, enhancing precision in navigating complex environments [25]. Dense direct SLAM methods, such as the RGB-iD SLAM system, utilize depth maps in inverse depth parametrization to enhance accuracy and efficiency in real-time localization [26], which is crucial for applications requiring precise mapping in dynamic environments.

The integration of neural fields with classical tracking and loop closure techniques represents a significant advancement, enabling real-time dense geometry and semantic segmentation [27]. Coarse-to-fine strategies, such as those in RLOCS, use image retrieval for initial pose estimation and observation constraints for refinement, improving localization accuracy in visually challenging scenarios [28]. Moreover, integrating semantic place categorization with traditional place recognition methods enhances robustness and accuracy, making systems resilient to environmental variations [29].

Reinforcement learning strategies have been employed to combine independent binary object-background segmentations, offering a flexible and modular approach to semantic segmentation

AI-generated, for reference only.

[9]. This method allows for adaptive learning of complex visual tasks, further enhancing robotic systems' perceptual capabilities.

Collectively, these advanced algorithms are instrumental in developing intelligent robotic systems capable of navigating and interacting with their environments with heightened accuracy and efficiency. Ongoing research significantly advances the field, particularly through integrating Open Knowledge models and foundational technologies. Recent developments have led to sophisticated vision models that recognize objects via language queries, effective navigation systems for mobile robots, and versatile grasping models capable of handling diverse objects. The OK-Robot framework exemplifies this progress, achieving a 58.5

1.4 Structure of the Survey

The survey is meticulously structured to provide a comprehensive examination of Semantic SLAM and 3D Mapping in robotics, beginning with an introduction that sets the stage for understanding the interdisciplinary nature and significance of intelligent robotic systems. A detailed background section delves into foundational concepts, defining key terms such as SLAM, Semantic Mapping, and Autonomous Navigation, while exploring the roles of Visual Perception and Computer Vision.

Subsequent sections explore recent technological advances in Semantic SLAM, highlighting the integration of semantic information, multi-sensor data, and real-time processing techniques. The survey transitions into examining various 3D mapping techniques employed in robotics, discussing the challenges of mapping dynamic environments and innovative data association methods.

The role of visual perception and computer vision is further explored in enhancing robotic navigation and mapping, focusing on techniques for visual place recognition, object detection, and the impact of deep learning. The discussion includes complexities of autonomous navigation within dynamic environments, examining critical aspects such as advanced path planning techniques, effective obstacle avoidance strategies, and the incorporation of semantic understanding through topological and semantic representations. This includes generating maps from natural language path instructions and employing deep perceptual feedback to enhance planning and responsiveness to human actions, thereby improving navigation accuracy and efficiency in previously unexplored settings [30, 31].

This study presents a variety of real-world applications and case studies illustrating the implementation of Semantic SLAM and 3D Mapping across multiple sectors, including autonomous vehicles, drones, and service robots, highlighting their roles in enhancing navigation, object recognition, and interaction capabilities in complex environments [32, 33, 11, 1, 16]. The survey concludes by identifying current challenges and future directions, discussing limitations in existing technologies and exploring potential areas for future research, thereby reinforcing the importance of ongoing innovation in this interdisciplinary field.The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 Foundational Concepts of SLAM and Semantic Mapping

Simultaneous Localization and Mapping (SLAM) is a fundamental technology in robotics, enabling the autonomous mapping of unknown environments while determining the robot's position. This dual capability is essential for effective interaction in complex and dynamic settings, where traditional navigation methods fall short. Traditional SLAM's reliance on geometric features often limits scene understanding and task-oriented navigation due to inadequate object representation [34]. Addressing issues like trajectory drift and inaccurate pose estimation necessitates advanced algorithmic strategies. The integration of multiple sensors, such as LiDAR and inertial encoders, significantly enhances SLAM accuracy, especially in environments unsuitable for single-sensor approaches [7].

Semantic Mapping extends SLAM by embedding semantic information into the mapping process, enhancing robotic systems' cognitive capabilities. This involves classifying and labeling environmental features, improving map usability and supporting high-level reasoning and navigation tasks [35]. Real-time integration of semantic labeling with geometric reconstruction is crucial for enhancing robotic cognitive functions [34]. Creating semantic maps that include points, planes, and objects as landmarks is vital for overcoming traditional SLAM limitations in dynamic environments [36].

AI-generated, for reference only.

Accurate data association is critical in SLAM, particularly in environments with similar or repetitive landmarks, complicating measurement assignments [35]. Existing SLAM methods often struggle with robust data association, leading to incorrect associations that bias results and cause algorithmic failures [35]. Perceptual aliasing from similar scenes can further result in incorrect inter-robot data associations, potentially causing failures in estimation back-ends [37].

Developing hierarchical metric-semantic maps in GPS-denied environments facilitates effective real-time communication and collaboration among robots [35]. Integrating RGB images and 3D point clouds for semantic segmentation and mapping offers potential solutions, although current methods often struggle to effectively combine geometric and semantic information [34].

Advancements in multi-modal semantic SLAM enable robots to distinguish between static and dynamic elements in complex environments, enhancing localization and mapping accuracy. The incorporation of semantic data augments cognitive capabilities, allowing robots to perform higher-level tasks, engage in language-based human-robot interactions, and navigate intelligently in real-world scenarios, even under challenging dynamic conditions [38, 16, 39, 11]. Ongoing research continues to address existing limitations and explore new frontiers in robotic perception and autonomy.

2.2 3D Mapping and Robotics

Integrating 3D mapping in robotics enhances the autonomy and functionality of mobile robotic systems. This process utilizes various sensing modalities, including LiDAR, stereo cameras, and inertial measurement units (IMUs), to construct detailed three-dimensional representations of environments. Such representations are crucial for tasks like obstacle detection, path planning, and efficient navigation, particularly in complex and dynamic settings [40]. LiDAR sensors offer advantages over cameras, being less affected by noise from distance and lighting variations, thus providing reliable data for robust mapping [41].

Recent advancements emphasize integrating geometric data with semantic segmentation from RGB cameras, resulting in a richer environmental understanding that supports informed decision-making [42]. This multimodal sensor fusion overcomes the limitations of individual sensors, enabling comprehensive environmental perception [43]. Techniques like the Fusion LiDAR-Inertial-Encoder SLAM (FLES) method exemplify this integration by combining data from multiple sources to achieve robust SLAM in texture-less environments, where traditional visual cues may be insufficient [6].

Efficient probabilistic 3D mapping frameworks, such as UFOMap, address inefficiencies in representing unknown spaces, reducing collision risks and improving data processing times [44]. Methods like Active Metric-Semantic SLAM (AM-SLAM) optimize exploration and localization in GPS-denied indoor environments using sparse semantic information, enhancing navigation and mapping capabilities in challenging conditions [45].

Innovative approaches, including parallel tracking-and-mapping with advanced depth estimation and pose recovery techniques, enhance visual odometry performance and overall mapping accuracy [4]. Integrating visual-inertial odometry with fiducial marker observations further refines mapping processes, providing a robust framework for precise localization [46].

3D mapping in robotics leverages diverse sensing modalities, such as 3D LiDAR and cameras, alongside advanced algorithmic strategies like SLAM and convolutional neural networks (CNNs). This multifaceted approach enhances robotic autonomy and functionality, facilitating high-level tasks across various applications, including navigation, surveillance, and virtual reality, by incorporating semantic information into 3D maps. Recent advancements, such as multimodal sensor-based systems and unified frameworks that improve compatibility across different LiDAR types, further enhance the accuracy and efficiency of 3D mapping in large-scale environments [32, 47]. Continuous advancements in this field are driving the development of more capable and adaptable robots, capable of navigating and interacting with their environments with unprecedented precision and understanding.

2.3 Role of Visual Perception in Robotic Systems

Visual perception is crucial in robotic systems, significantly enhancing their ability to interpret and interact with their environments. Integrating semantic segmentation into mapping processes allows for differentiating between static and dynamic objects, improving situational awareness and decision-making capabilities [48]. This capability is vital in dynamic environments, where traditional visual SLAM systems may struggle with accuracy and robustness due to moving objects [49].

The dynamic nature of environments presents challenges in representing and learning affordances compared to static objects [50]. To address these challenges, systems like TwistSLAM++ have been developed, enhancing object tracking and pose estimation by fusing stereo images and LiDAR information, thereby improving SLAM robustness in dynamic settings [51].

Visual perception is integral to localization within robotic systems. The Sparse Feature Pyramid (SFP) method employs varying descriptor lengths for keypoints, enhancing localization accuracy by effectively capturing essential visual features [17]. This is complemented by techniques such as the Probabilistic Egocentric Motion Correction and Projection (PEMCP), which corrects lidar point cloud data for ego-motion distortion and projects it onto camera images while incorporating uncertainty estimates, refining perception and mapping processes [52].

Incorporating semantic information into perception systems fosters more informed reactive planning, enhancing the robot's understanding of its environment and enabling more effective navigation strategies [31]. This is significant in partially observable environments, where mapping or localization uncertainty can limit the effectiveness of existing control synthesis methods [53].

Innovative approaches such as ACUMEN integrate active exploration, open set recognition, and the generation of human-understandable explanations, distinguishing them from traditional methods and highlighting the evolving role of visual perception in enhancing robotic autonomy and human-robot interaction [54].

The role of visual perception in robotic systems is multifaceted and essential, as it enhances mapping and localization accuracy, improves handling of dynamic objects, and facilitates robust visual relocalization, particularly in challenging environments where traditional methods may falter. This includes advancements in visual place recognition for aerial imagery, optimized feature selection using spatial attention, and the integration of multi-camera systems for comprehensive environmental awareness, all contributing to the overall effectiveness and reliability of robotic navigation and interaction [55, 56, 25, 57, 58]. These advancements are vital for developing intelligent robotic systems capable of operating autonomously in complex and dynamic environments.

3 Technological Advances in Semantic SLAM

3.1 Integration of Semantic Information

Integrating semantic information into SLAM systems greatly enhances robotic cognitive and operational capabilities, facilitating precise navigation and interaction in complex environments. Techniques such as semantic feature embedding within the 3D Gaussian framework offer improved scene interpretation over traditional color-based methods [34]. Systems like Kimera-Multi demonstrate the efficacy of distributed SLAM, with its fully distributed pose graph optimization surpassing existing methods in robustness and processing [35]. The Adaptive Navigation Scheme exemplifies advancements by incorporating diverse visual odometry modalities and a two-stage navigation scheme, improving localization accuracy [7].

Active Semantic Loop Closure (SLC) maintains map accuracy by dynamically balancing exploration and uncertainty reduction [45]. Multi-modal SLAM frameworks, such as LOCUS, integrate multi-stage scan matching and health-aware sensor integration, enabling robust sensor fusion and failure handling [59]. Datasets like the Oxford Multimotion Dataset and SeePerSea are crucial for developing multimotion estimation algorithms and robust maritime autonomy, respectively, highlighting the importance of comprehensive datasets for advancing semantic integration [37, 42].

Innovations like Vision-Language Models (VLMs) and Large Language Models (LLMs) significantly enhance robotic capabilities in object recognition, navigation, and manipulation, as demonstrated by the OK-Robot's success in complex tasks. The application of foundation models in robot learning further improves environmental interaction and decision-making, paving the way for versatile robotic applications in dynamic real-world settings [60, 61, 23, 20, 62].

3.2 Integration of Multi-Sensor Data

The integration of multi-sensor data in SLAM systems is vital for enhancing the robustness and precision of robotic perception and navigation. By leveraging data from cameras, LiDAR, and IMUs, these systems overcome the limitations of single-sensor approaches. TwistSLAM++ exemplifies this by combining stereo images and LiDAR data to effectively track moving objects and accurately estimate their poses, enhancing SLAM robustness in dynamic environments [51]. Datasets like the Oxford Multimotion Dataset provide essential resources for developing and testing multi-sensor SLAM algorithms, offering comprehensive data for benchmarking and refining integration techniques [37].

Advanced SLAM frameworks utilize hierarchical structures and Gaussian primitives for efficient data management, crucial for real-time applications. Techniques like LiDAR-guided Image Fusion (LI-Fusion) enhance point cloud features by integrating semantic information from images, improving 3D object detection accuracy [52, 63, 47]. Challenges in multi-sensor integration, such as data synchronization and depth ambiguity, are addressed by innovative solutions like multi-task learning approaches, enhancing depth completion and 3D mapping accuracy [14, 64].

The integration of multi-sensor data marks a significant advancement in robotics, improving environmental perception and navigation by combining various sensing modalities. This approach enhances landmark recognition accuracy and data association in complex urban environments, addressing noise and uncertainty challenges in sensor data. Employing modular architectures and probabilistic models, modern SLAM systems achieve robust performance in dynamic settings, leading to improved navigation efficiency and reduced localization drift [33, 12, 13].

3.3 Real-Time Processing and Optimization

Real-time processing and optimization are crucial for advancing Semantic SLAM systems, enabling efficient navigation and decision-making in dynamic environments. Probabilistic data association methods, such as the k-best assignments approach, enhance real-time processing by efficiently computing assignment probabilities, ensuring accuracy in challenging conditions [65]. Techniques like Kinematic-ICP refine odometry estimates by integrating LiDAR data with kinematic constraints, producing smooth and accurate motion estimates [66].

Methods like SemGauss-SLAM, using 3D Gaussian representations, enhance real-time processing capabilities for accurate semantic mapping and robust tracking [34]. The AM-SLAM approach leverages semantic information to improve localization accuracy while managing exploration tasks efficiently [45]. Advanced frameworks such as Kimera-Multi employ robust distributed optimization techniques to mitigate outlier loop closures, ensuring accurate trajectory estimation and enabling independent navigation with local communication for loop closure detection [35]. SlideSLAM demonstrates the effectiveness of distributed processing in real-time SLAM by utilizing sparse semantic object representations for enhanced localization accuracy and map merging efficiency [6].

Integrating learning and planning in active 3D mapping significantly improves map accuracy, outperforming state-of-the-art approaches by enhancing recall and efficiency [67]. This indicates the potential of combining machine learning techniques with real-time optimization strategies to advance Semantic SLAM capabilities. Ongoing advancements in Open Knowledge Models, Large Language Models, and novel planning methods are crucial for enhancing autonomous robotics, enabling adaptive systems capable of effective object recognition, navigation, and manipulation. The OK-Robot framework's success in open-ended tasks underscores the importance of nuanced details in combining robotic modules. The shift toward foundation models in robot learning highlights the potential for real-world applications across manipulation, navigation, planning, and reasoning, emphasizing the need for research on multimodal interactions and AI alignment to further enhance robotic capabilities [20, 61, 23].

In recent years, the advancement of robotics has been significantly influenced by the development of sophisticated 3D mapping techniques. These techniques are crucial for enabling robots to navigate and interpret their surroundings effectively. As illustrated in Figure 2, the hierarchical structure of these mapping techniques can be categorized into three main areas: handling dynamic environments, innovative data association techniques, and advanced mapping and localization. Each of these categories is further subdivided into specific challenges, innovations, and applications. This classification not only emphasizes the complexity of the field but also highlights the integration of semantic information, multimodal data, and advanced algorithms. Such integration is essential for enhancing the precision and adaptability of robotic systems in complex environments, ultimately contributing to their operational efficiency and effectiveness.

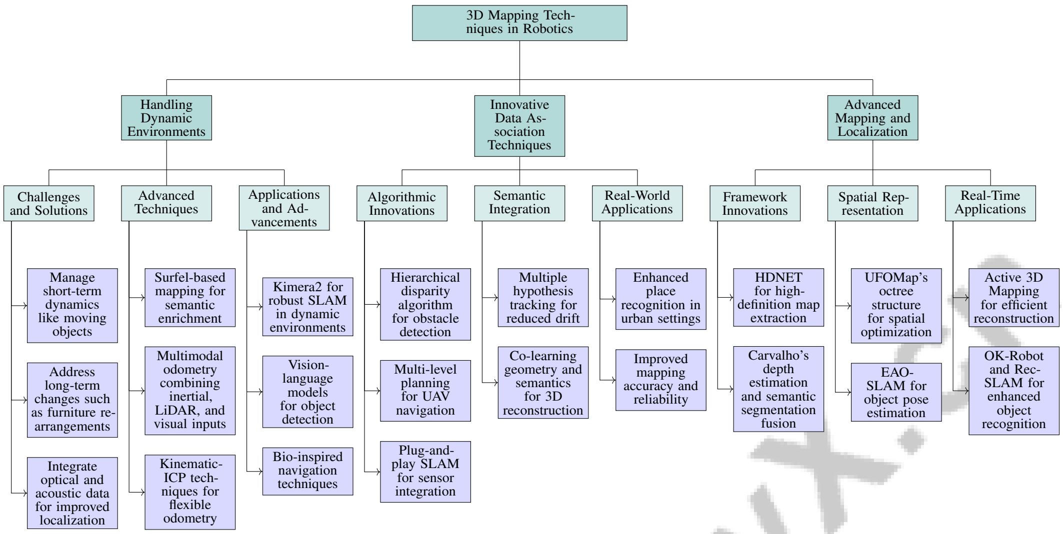


Figure 2: This figure illustrates the hierarchical structure of 3D mapping techniques in robotics, categorized into handling dynamic environments, innovative data association techniques, and advanced mapping and localization. Each category is further subdivided into specific challenges, innovations, and applications, highlighting the integration of semantic information, multimodal data, and advanced algorithms to enhance robotic system precision and adaptability in complex environments.

4 3D Mapping Techniques in Robotics

4.1 Handling Dynamic Environments

Addressing the challenges of mapping dynamic environments is pivotal for deploying robotic systems in real-world applications. This involves managing both short-term dynamics, like moving objects, and long-term changes, such as furniture rearrangements, which can affect mapping accuracy. Integrating optical and acoustic data enhances localization in dynamic settings without needing prior object class knowledge, thereby improving adaptability [68]. Surfel-based mapping techniques effectively navigate dynamic environments by merging semantic information with 3D laser scans, thus improving mapping accuracy and efficiency. They address issues like feature sparsity and motion blur, creating semantically enriched maps that filter out moving objects and enhance localization through semantic constraints. These methods significantly outperform traditional geometric approaches, especially in environments with substantial dynamic elements [48, 69, 38].

Multimodal odometry and mapping methods that combine inertial data with LiDAR and visual inputs have proven effective in dynamic environments, such as those encountered by rail vehicles [70]. This integration ensures accurate odometry and mapping, addressing the challenges posed by dynamic conditions. Kinematic-ICP techniques further enhance robustness by dynamically adjusting the weighting between LiDAR and wheel odometry, offering flexibility across diverse scenarios [66]. Additionally, leveraging multi-frame semantic associations to establish constraints significantly reduces cumulative drift, improving mapping precision [34].

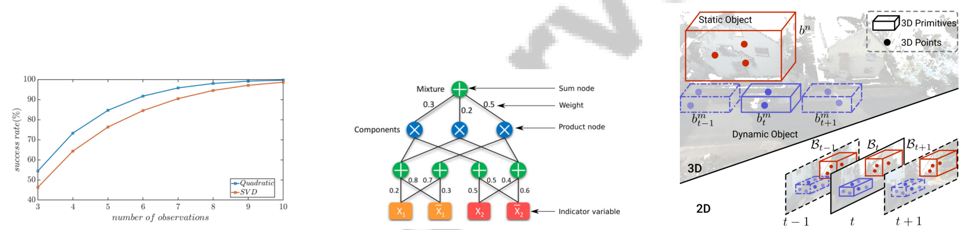
Experiments on diverse datasets from real and simulated robotic platforms, including drones and wheeled robots, underscore the effectiveness of advanced SLAM methods like Kimera2 in handling dynamic environments [18]. These methods consistently outperform state-of-the-art SLAM techniques, maintaining mapping accuracy amidst environmental changes. A combination of advanced algorithmic strategies, robust sensor integration, and innovative frameworks is essential for effectively handling dynamic environments. Recent advancements, including vision-language models for object detection and bio-inspired navigation techniques, significantly enhance the capabilities of autonomous systems, enabling reliable and efficient operations in complex settings such as intricate pick-and-drop tasks in homes, inspections in hazardous areas, and navigation of unknown terrains with enhanced safety and speed [20, 71, 72, 60].

4.2 Innovative Data Association Techniques

Innovative data association techniques are crucial for enhancing the accuracy and efficiency of 3D mapping in robotics, especially within complex and dynamic environments. Keller et al.'s hierarchical disparity algorithm exemplifies this innovation, utilizing a trinocular camera setup and semantic object triangulation to detect thin obstacles and reflective surfaces that traditional SLAM systems struggle to manage [73]. This method improves obstacle detection and the overall robustness of the mapping process by integrating semantic triangulation with disparity estimation.

In autonomous flight, Liu et al. developed a multi-level planning and mapping framework that computes feasible trajectories for UAVs, enhancing navigation in complex environments and optimizing data association by incorporating semantic information for effective real-time operations [74]. Colosi et al. introduced a plug-and-play SLAM system that integrates various sensor types into a unified framework, enhancing flexibility and performance crucial for effective data association in diverse environments [12]. This seamless integration allows the system to leverage complementary data, improving mapping accuracy and reliability.

Recent studies demonstrate substantial advancements in 3D mapping through the integration of semantic information and multiple hypothesis tracking. These methods effectively tackle the complexities of real-world environments, such as urban settings with visually similar surroundings, by reducing measurement ambiguity and enhancing place recognition. Utilizing multiple hypothesis trees creates a probabilistic framework for semantic measurements, achieving a 33



(a) The graph compares the success rates of two methods, Quadratic and SVD, as the number of observations increases.[75]

(b) A diagram illustrating a mixture model with three components[76]

(c) Dynamic Object Segmentation in 3D Images Using 2D Primitives[14]

Figure 3: Examples of Innovative Data Association Techniques

As illustrated in Figure 3, advancements in 3D mapping techniques have greatly improved robotics' interpretation and interaction capabilities within complex environments. Innovative data association techniques are integral to effectively integrating and interpreting sensory data. The examples provided showcase three distinct approaches: a comparative analysis of Quadratic and Singular Value Decomposition (SVD) methods, highlighting their effectiveness as observation numbers increase; a mixture model diagram illustrating the decomposition of complex data; and dynamic object segmentation in 3D images using 2D primitives, demonstrating how simplifying assumptions can enhance real-time object detection and tracking accuracy. Collectively, these examples underscore significant strides made in data association techniques, crucial for the evolution of autonomous robotic systems.

4.3 Advanced Mapping and Localization

Advanced mapping and localization techniques are fundamental to improving the precision and efficiency of autonomous robotic systems in complex environments. The HDNET framework exemplifies innovation by leveraging high-definition maps to extract geometric and semantic features, incorporating a map prediction module to estimate maps from LiDAR data in real-time, thus enhancing 3D object detection and environmental understanding [77]. Carvalho et al. propose a method that fuses depth estimation and semantic segmentation within a single framework, improving mapping cohesion and accuracy by addressing traditional technique limitations [2].

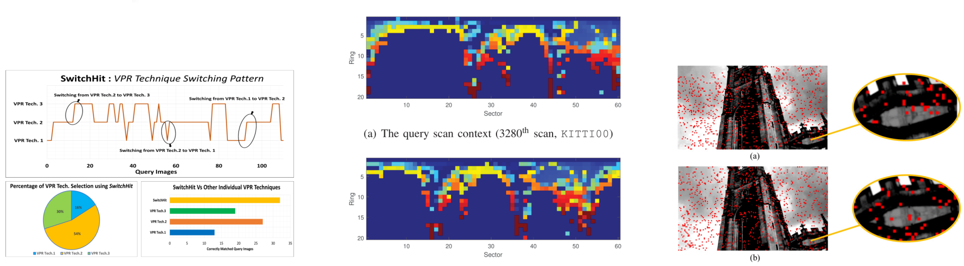
UFOMap represents a significant advancement, utilizing an octree structure to explicitly delineate occupied, free, and unknown spaces, optimizing spatial information representation, reducing computational overhead, and enhancing mapping accuracy [44]. EAO-SLAM explores advanced mapping and localization methods by employing robust algorithms for object pose estimation, optimizing camera and object poses to ensure alignment critical for maintaining consistency and precision in dynamic environments [78]. Active 3D Mapping, as defined by Zimmermann et al., iteratively reconstructs dense occupancy maps from sparse measurements, optimizing depth-measuring ray selection for efficient and accurate map reconstruction essential for real-time robotic applications [67].

These advanced techniques represent significant progress in robotics, offering robust solutions for navigating complex environments. By leveraging innovative frameworks like the Open Knowledge-based robotics system (OK-Robot), which integrates Vision-Language Models for object recognition and navigation, along with edge-computing techniques for multi-robot simultaneous localization and mapping (RecSLAM), these methods enhance robotic capabilities. This integration not only improves performance in tasks such as open-ended pick-and-drop operations—achieving a 58.5

5 Visual Perception and Computer Vision

5.1 Techniques for Visual Place Recognition

Visual Place Recognition (VPR) is critical for robotic systems, enabling them to identify and recollect specific locations essential for navigation and autonomous functionality. A primary challenge in VPR is maintaining reliable recognition despite variations in viewpoint, lighting, and environmental conditions. Advanced techniques often employ multi-camera systems to enhance perception; for instance, integrating multiple fisheye cameras improves visual place recognition in self-driving cars by providing a comprehensive view that aids in precise localization [55]. Semantic information integration further boosts VPR performance by offering a higher-level scene understanding. Systems like LOCUS utilize LiDAR data to capture geometric and temporal aspects, enhancing place recognition [79]. Semantic localization frameworks enhance robust place recognition by integrating keypoint detectors and feature extractors across diverse environments [80]. Event-based stereo visual odometry systems use stereo configurations of event cameras to improve depth perception and performance in dynamic lighting conditions [4]. SQ-SLAM exemplifies the integration of semantic information with SLAM systems, offering accurate object shape representation and employing lightweight data association for real-time performance [81]. Object-based state estimators like OBM-Net sequentially process observations to predict object properties over time, contributing to enhanced place recognition [82]. The combination of models like YOLOv5 for images and PointPillar for point clouds underscores the effectiveness of merging image and point cloud data [42]. The exploration of advanced VPR techniques continues to drive the development of intelligent robotic systems capable of operating in complex environments. Innovations in semantic place categorization and probabilistic switching mechanisms enhance robustness against environmental variations, facilitating precise location identification and supporting efficient robotic operations across various platforms and real-world scenarios [57, 58, 29].



(a) SwitchHit: VPR Technique Switching Pattern[57]

(b) The query scan context (3280th scan, KITTI00)[83]

(c) Deep Learning in Image Segmentation: A Comparative Study[25]

Figure 4: Examples of Techniques for Visual Place Recognition

As shown in Figure 4, various techniques in visual perception and computer vision address the challenges in VPR. The "SwitchHit: VPR Technique Switching Pattern" depicts a dynamic selection process among multiple VPR techniques, emphasizing method adaptability based on query image counts. The "query scan context" within the KITTI00 dataset visually represents the likelihood of a query's presence in specific scan regions, underscoring the importance of spatial awareness for accurate place recognition. Lastly, "Deep Learning in Image Segmentation: A Comparative Study" showcases deep learning algorithms' effectiveness in segmenting images, comparing methods in recognizing features such as a church tower. These techniques collectively exemplify diverse strategies in visual perception and computer vision, contributing to robust place recognition in complex environments [57, 83, 25].

5.2 Object Detection and Scene Understanding

Object detection and scene understanding are vital for autonomous robotic systems, enabling precise perception and interpretation of environments. Advanced methods enhance these capabilities through multi-sensor data fusion and semantic information integration. The EPNet framework effectively combines multi-sensor data without requiring image annotations, enhancing detection accuracy through confidence consistency [63]. The Embodied Semantic Fovea system integrates RGB-D images and gaze data to create detailed 3D representations, improving object detection and scene understanding [84]. This comprehensive environmental data capture enhances the robot's navigation and interaction capabilities. Incorporating reinforcement learning strategies into semantic segmentation processes increases flexibility for handling various tasks and learning from fewer labels, thereby boosting segmentation performance [9]. This adaptability allows robotic systems to function efficiently in new environments with minimal supervision. The fusion of UAV and UGV sensor data, as proposed by Surmann et al., utilizes a novel registration method to create accurate 3D maps, facilitating effective localization and scene understanding [85]. This integration enables robotic systems to operate effectively across diverse environments by leveraging different sensor modalities. Advancements in object detection and scene understanding are essential for developing intelligent robotic systems capable of autonomous operation in complex environments. Techniques like semantic segmentation and multi-task learning enhance robots' ability to classify and interpret scenes, informing them about the global state of their surroundings. The integration of vision-language models in frameworks like OK-Robot highlights the importance of nuanced details for effective navigation, object recognition, and manipulation, ultimately improving robotic performance in real-world applications [20, 64, 24, 54]. These innovations continue to propel the evolution of robotics, paving the way for more efficient and scalable operations.

5.3 Impact of Deep Learning on Visual Perception

Deep learning has revolutionized visual perception in robotics, significantly enhancing robotic systems' ability to accurately and efficiently interpret and interact with their environments. The integration of deep learning techniques into visual odometry and mapping has been crucial for maintaining consistent state estimates, particularly in GPS-denied environments. DLIOM exemplifies this integration, demonstrating substantial improvements in localization accuracy and mapping resiliency compared to traditional methods, showcasing its effectiveness in real-world applications [86]. Multiple hypothesis semantic mapping frameworks have further advanced visual perception by enhancing robustness to measurement noise and ambiguity. These frameworks efficiently manage computational complexity through hypothesis management, allowing robotic systems to navigate complex environments with reduced error rates [33]. This is particularly beneficial in dynamic settings where traditional methods may struggle with data association and noise. The LOCUS framework highlights the adaptability and robustness of deep learning-enhanced systems, providing high accuracy in challenging environments and effectively handling sensor failures [59]. The ability to adapt across different robotic platforms further emphasizes the versatility of deep learning methodologies in enhancing visual perception capabilities. The impact of deep learning on visual perception in robotics is profound, driving advancements in accuracy, efficiency, and adaptability. Recent innovations in autonomous robotics, such as vision-language models, autonomous planning, and edge computing, significantly enhance robotic capabilities. For instance, the OK-Robot framework integrates object recognition, navigation, and manipulation without prior training, achieving a 58.5

6 Autonomous Navigation in Dynamic Environments

6.1 Path Planning and Obstacle Avoidance

Path planning and obstacle avoidance are integral to autonomous robotics, enabling efficient navigation in complex environments. Addressing goal-oriented motion, the DroNet framework exemplifies the need for goal-directed strategies [87]. The Area Graph method enhances path planning by simplifying environmental representation, thus mitigating over-segmentation [88]. In uncertain obstacle detection, Multiple-Hypothesis Path Planning (MHPP) constructs probabilistic graphs to facilitate decision-making in unpredictable scenarios [89]. LiDAR-based systems, like those for over-canopy navigation, demonstrate autonomy without GPS reliance, ensuring operation in GPS-denied settings [90].

The SC-Explorer method improves path planning by integrating measured and predicted data, allowing adaptive strategies responsive to environmental changes [71]. Incorporating height and slope data in voxel map-based algorithms enables precise obstacle avoidance for both aerial and ground navigation [91]. In specialized applications, such as robotic retinal surgery, trajectory planning ensures precision in reaching targets [19]. The LiDAR Road-Atlas refines urban path planning by addressing dynamic objects and defining traversable areas [40].

End-to-end training in simulation enhances navigation by minimizing the sim2real gap, facilitating effective transitions to real-world environments [36]. Recent advancements, including multiple-hypothesis path planning, have significantly improved navigation success rates in challenging scenarios, empowering robots to navigate dynamic environments with goal-oriented motion [87, 92, 89].

6.2 Semantic Understanding and Mapping

Semantic understanding and mapping are vital for autonomous navigation in dynamic environments, enhancing localization and loop closure detection [93]. The SlideSLAM approach utilizes lightweight representation to facilitate efficient inter-robot communication and real-time operations [6]. Integrating semantic information into SLAM systems enhances decision-making, as seen in the STAM framework, which evaluates dynamic interactions for behavioral adaptation [50].

In underwater environments, integrating visual odometry with probabilistic mapping improves navigation accuracy [7]. Such techniques are crucial where traditional methods fall short. Semantic understanding and mapping advancements, including Vision-Language Models (VLMs) and probabilistic generative models, significantly enhance robotic intelligence and adaptability, enabling complex tasks like open-ended pick-and-drop operations with high success rates [20, 61, 23].

6.3 Challenges and Solutions in Dynamic Environments

Navigating dynamic environments presents challenges such as constrained perception due to sensor noise and GPS dependence, which can fail under dense canopies [90]. Sensor performance degradation in repetitive environments complicates localization, leading to errors [94]. Advanced architectures like DroNet, combined with SLAM and Dijkstra's path planning, address goal-oriented motion in dynamic settings [87]. However, methods relying on distinct features, such as those used in glacial environments, face limitations due to feature scarcity [95].

Surfel-based navigation techniques enhance performance in unstructured environments but may struggle with rapid changes affecting prediction accuracy [96]. SLAM systems' performance can be limited by motion range and scene complexity, necessitating adaptable solutions [97]. Collaborative UAV mapping enhances exploration efficiency, reducing collision rates, though reliance on visual data can introduce inaccuracies in rapidly changing environments [98, 99].

Future research should focus on integrating vision-based global localization and 3D semantic scene parsing to bolster navigation capabilities. Developing datasets emphasizing challenging conditions will provide benchmarks for evaluating algorithms under dynamic variations, requiring continuous innovation and robust algorithm integration to enhance robotic autonomy in dynamic environments [92, 100].

7 Applications and Case Studies

The integration of advanced navigation and mapping technologies has revolutionized robotics, particularly in deploying autonomous systems within complex environments where traditional methods fall short. This section examines specific applications, highlighting Semantic SLAM's transformative role in autonomous medical robots and its impact on healthcare.

7.1 Autonomous Medical Robots

Semantic SLAM represents a substantial advancement in autonomous medical robots, enabling precise navigation and task execution in complex hospital settings. These robots leverage deep learning, SLAM, and 3D mapping for efficient movement through dynamic environments, performing tasks such as supply delivery and patient monitoring while adeptly avoiding obstacles [87, 92, 72, 23]. Semantic SLAM facilitates detailed semantic maps crucial for effective operations in medical settings. A significant challenge is accurate object association and pose estimation, essential for interacting with medical equipment. Advanced techniques improve object association, enhancing task execution and minimizing errors [101].

Multi-sensor data integration, including RGB-D cameras and 3D LiDAR, enhances perception capabilities, improving mapping and navigation accuracy in clinical settings [10, 97, 32]. Adapting to environmental changes, such as moving patients and staff, is critical for ensuring safety and efficiency. The integration of Semantic SLAM technology offers transformative potential, particularly in reducing cross-infection risks through contactless operations, enhancing routine task efficiency amid pandemic strains on medical systems. Ongoing research into semantic object association and real-time pose refinement is essential for accurate tracking and interaction, allowing healthcare professionals to focus on critical patient care, ultimately improving outcomes [10, 101].

7.2 Service Robots in Dynamic Environments

Service robots in dynamic environments face unique challenges requiring advanced perception and navigation capabilities. These robots must navigate intricate, unpredictable settings such as crowded public spaces and dynamic industrial facilities, where rapid environmental changes demand robust visual place recognition and real-time decision-making [8, 29, 87, 57, 92]. Semantic SLAM enhances adaptability and efficiency in these settings.

Multi-sensor data fusion, integrating cameras, LiDAR, and other sensors, provides a comprehensive environmental understanding, crucial for accurate mapping and obstacle avoidance [51]. Innovative data association techniques ensure effective navigation despite environmental changes [73]. Advanced algorithms incorporating semantic information enhance visual perception, enabling recognition and response to dynamic elements [4].

Robust real-time processing and optimization strategies, such as probabilistic data association and kinematic-ICP, refine localization and mapping, enabling navigation strategy adjustments in response to environmental changes [66]. These strategies ensure optimal performance and safety in complex settings.

Advanced perception systems, SLAM technologies, and deep learning approaches enable service robots to navigate safely around obstacles and execute complex tasks autonomously, optimizing performance in applications like home tidying and retail management [8, 87, 23, 20, 102]. These innovations drive more intelligent and adaptable robotic systems, enhancing efficacy in diverse service contexts.

7.3 Precision Farming with LiDAR-based Navigation

LiDAR-based navigation in precision farming represents a significant technological advancement, enabling autonomous vehicles to operate with precision and efficiency. These systems excel in detecting and navigating crop rows, crucial for precision farming operations requiring accurate alignment and path following [90]. Independent of GPS, they are invaluable in environments where GPS signals are unreliable, such as dense canopies or remote areas.

LiDAR sensors create detailed 3D maps, facilitating precise obstacle detection and avoidance, essential for the safe operation of autonomous farming equipment. Integrating LiDAR-based navigation with advanced control algorithms enhances trajectory precision in dynamic environments through semantic information incorporation [52, 38, 66]. This integration promotes site-specific management, optimizing resource inputs and maximizing crop yields by accommodating diverse scenarios and conditions [90, 47, 103, 104].

LiDAR-based navigation optimizes field management, increases yield precision, and reduces reliance on traditional methods. Innovations like AgriColMap enhance agricultural robots' mapping capabilities, integrating data from aerial and ground vehicles for accurate field representations. Incorporating semantic information into mapping processes further optimizes autonomous farming equipment, driving sustainable and productive farming practices [103, 90, 38, 47, 104].

7.4 3D Perception and Mapping in GPS-denied Environments

3D perception and mapping in GPS-denied environments necessitate innovative solutions due to traditional navigation systems' limitations. Recent advancements in LiDAR-based SLAM frameworks, such as the extended Cartographer SLAM library, enhance compatibility across diverse sensing payloads, enabling real-time high-quality 3D map creation. Visual geolocalization techniques address GNSS shortcomings, adapting Visual Place Recognition methods to overcome environmental variations [58, 47, 105]. These environments, often lacking satellite signals, require alternative mapping and navigation methods.

Integrating LiDAR with visual and inertial sensors enhances perception capabilities in GPS-denied environments, achieving high precision in mapping and localization. Techniques like the Fusion LiDAR-Inertial-Encoder SLAM method exemplify robust SLAM solutions where traditional visual cues are inadequate [6]. Probabilistic egocentric motion correction and projection methods preserve uncertainty estimates, enhancing real-time SLAM robustness [52].

Frameworks like Kimera-Multi utilize distributed optimization to minimize outlier loop closures' impact, ensuring accurate trajectory estimation and enabling independent navigation with local communication for loop closure detection [35]. Hierarchical metric-semantic maps are crucial for effective communication and collaboration in GPS-denied settings [35].

Advancements in 3D perception and mapping techniques are essential for developing autonomous systems, enabling navigation in complex settings like underwater and aerial environments, where traditional methods are hindered by limited visibility and signal instability. Innovations in sensor integration, algorithmic strategies, and distributed processing foster adaptable and intelligent robotic systems, expanding applicability across various domains [64, 47, 91, 58, 106].

7.5 Collaborative UAV Missions in Emergency Scenarios

Collaborative UAV missions in emergency scenarios highlight Semantic SLAM's critical application, enhancing rapid response and situational awareness in complex environments. These missions leverage multiple UAVs' collaborative capabilities, improving data collection efficiency and coverage. Techniques like Visual SLAM and Structure-from-Motion enable real-time habitat mapping and monitoring, facilitating effective conservation efforts. Sensor data integration generates denser 3D models, enhancing accuracy and resource efficiency [64, 99]. Semantic SLAM facilitates semantically rich map creation, essential for decision-making and coordination during emergencies.

A primary challenge is ensuring accurate communication between UAVs for synchronization and data consistency. Decentralized SLAM approaches enable efficient information sharing, enhancing collaboration and adaptability to changing conditions [6]. Operating in GPS-denied environments is another advantage, with multi-sensor data fusion maintaining accurate localization and mapping [35].

Advanced data association techniques improve object detection and scene understanding accuracy, enabling UAVs to distinguish between static and dynamic elements for informed decision-making [4]. Integrating Semantic SLAM enhances emergency response operations' efficiency and effectiveness, supporting dynamic object recognition and localization. UAVs can execute missions with greater accuracy and safety, providing critical support in emergencies where rapid situational awareness is essential [11, 74, 99]. Ongoing advancements continue to propel more intelligent and capable UAV systems, paving the way for future innovations in emergency management and disaster relief.

8 Challenges and Future Directions

In robotics, addressing the challenges of Simultaneous Localization and Mapping (SLAM) technologies is crucial for advancing autonomous systems. Understanding current SLAM limitations is vital for identifying potential improvements and innovations. This section explores specific limitations that affect SLAM effectiveness across various environments, laying the groundwork for future research and development.

8.1 Limitations in Current SLAM Technologies

Current SLAM technologies face several challenges limiting their effectiveness in diverse environments. A major issue is the dependency on high-quality input data, which restricts their use in feature-poor or sensor-degraded conditions [52]. Computational demands of methods like LiDAR Road-Atlas impede real-time application due to intensive normal computation and nearest neighbor searches [66]. LOCUS, while robust, has a higher CPU load compared to other methods, limiting its deployment on less powerful platforms [59].

Variability in object appearance and sparse semantic information present significant challenges, as probabilistic methods struggle in highly variable environments [34]. The need for extensive training data further limits object-based state estimators' applicability [9]. Enhancing dense RGB-D semantic mapping through additional sensor modalities could improve performance across diverse conditions [35].

Despite supporting various sensors, plug-and-play SLAM systems' performance may vary based on configurations and environmental conditions [42]. Probabilistic data association methods face scalability issues with high-dimensional assignment problems [37]. Additionally, reliance on initial sparse measurements can lead to inaccuracies in dynamic environments [7].

Decentralized SLAM methods' abstraction of raw sensor data to sparse semantic landmarks can cause information loss, affecting map accuracy [50]. High-quality ego-motion estimation remains a critical limitation, influenced by environmental factors [52]. Methods dependent on accurate robot dynamics modeling may not capture all real-world variabilities, affecting performance in unforeseen scenarios [36]. These limitations highlight the need for continued SLAM innovation, focusing on sensor integration, computational efficiency, and adaptability.

8.2 Integration of Deep Learning and SLAM

Integrating deep learning with SLAM systems marks a pivotal advancement in enhancing autonomous robotic systems' perceptual and navigational capabilities. Deep learning, particularly in feature detection and place recognition, significantly improves SLAM systems' accuracy and robustness. Future research should focus on integrating wind measurements into SLAM models, exploring predictive inference for informative path planning, and developing multi-agent systems for collaborative mapping in complex environments [107].

A primary challenge in this integration is the reliance on extensive training data, often impractical in diverse real-world scenarios. Addressing this is crucial for advancing deep learning and SLAM integration. Future research should also incorporate loop closure and sliding window optimization techniques to reduce drift and enhance performance in dynamic environments [108].

Refining deep learning integration for dynamic environments and exploring additional sensor integrations can enhance performance [18]. Developing multimodal sensor-based semantic 3D mapping approaches, especially for indoor environments, is promising. Synergizing SLAM and semantic processing could significantly improve robotic systems' cognitive capabilities.

Enhancing CNN models for image retrieval and semantic segmentation, along with incorporating additional geometric constraints, could improve SLAM localization accuracy. Integrating multi-camera systems and low-cost sensors improves localization accuracy and reliability. A surround multi-camera setup eliminates blind spots and enhances environmental perception, providing comprehensive visual data for autonomous navigation and obstacle detection applications. This approach addresses GNSS limitations, which suffer from signal instability, while leveraging techniques like semantic edge alignment and spatial attention for optimized feature selection, resulting in precise localization outcomes in diverse conditions [55, 109, 110, 25, 58].

Incorporating adaptive resource allocation capabilities, such as those exemplified by Dynamic Resource Allocation for Autonomous Applications (DRAA), can significantly enhance deep learning and SLAM integration. These capabilities optimize computational resources, crucial for real-time dense computer vision and SLAM systems. Frameworks like SLAMBench, which benchmark performance, accuracy, and energy consumption, illustrate adaptive techniques' potential to improve operational efficiency and reliability in complex environments. This is particularly relevant for autonomous systems relying on robust 3D object detection and mapping, as seen in frameworks leveraging High-Definition maps and neural implicit representations for superior performance [111, 86, 77, 112]. Future research could focus on developing dynamic semantic memory systems and improving human-robot interaction to enhance task success rates. Transformer-based approaches and collaborative perception strategies utilizing multi-agent systems for improved data sharing and processing are promising areas for further exploration.

The integration of deep learning with SLAM systems offers significant potential for advancing robotics, enabling more intelligent, adaptable, and efficient autonomous systems. Continued research and innovation in robotic perception and navigation are crucial for addressing existing limitations and exploring new opportunities, particularly through the integration of advanced technologies such as Open Knowledge-based frameworks, Vision-Language Models, and foundation models. These developments have already demonstrated significant improvements in robotic performance, such as the 58.5

8.3 Advancements in AI and Machine Learning

Recent advancements in AI and machine learning have significantly impacted Semantic SLAM systems, enhancing their adaptability, computational efficiency, and robustness in dynamic environments. The integration of edge computing, as demonstrated by RecSLAM, has notably reduced processing latency by up to 39.31% compared to traditional cloud-based SLAM solutions, which is critical for real-time applications [22]. This reduction in latency is essential for ensuring the effectiveness of SLAM systems in real-time scenarios.

Exploring advanced deep learning techniques for depth estimation and semantic labeling shows potential for improving 3D reconstruction quality. Future research could focus on refining these techniques to enhance SLAM initialization procedures and overall accuracy [113]. Additionally, integrating inertial measurements into event-based stereo visual odometry systems may improve motion estimation, providing more reliable temporal correspondence in event data [4].

Incorporating additional sensor modalities, such as color data from cameras on UGVs, presents an opportunity to improve correspondence search and mapping accuracy in multi-hybrid systems [85]. Further research could explore optimizations for real-time processing and robustness, particularly in highly dynamic environments where traditional methods may struggle.

The development of plug-and-play SLAM systems offers flexibility and user-friendliness, with future research potentially focusing on specific sensor optimizations and intuitive configuration tools to enhance user experience [12]. Advancements in semantic feature extraction techniques, as explored in SemGauss-SLAM, could further improve system performance in complex and dynamic environments [34].

Future research could also enhance methods like Kinematic-ICP to better handle non-planar environments and optimize adaptive regularization strategies [66]. Improved motion estimation techniques that address occlusions and leverage datasets with diverse scenarios are another area for exploration [37]. Enhancing loop closure detection modules and testing systems in scenarios with intermittent communication are crucial for improving distributed SLAM frameworks [35].

Furthermore, future research will explore closed-loop adaptation of dynamics, improved noise estimation, and the integration of additional sensory modalities to enhance navigation capabilities [36]. Continuous advancements in AI and machine learning are driving the evolution of Semantic SLAM, enabling more intelligent, adaptable, and efficient autonomous systems capable of navigating complex and dynamic environments. These innovations are crucial for expanding the frontiers of robotic perception and navigation, paving the way for future breakthroughs in autonomous systems.

8.4 Future Research Directions in Semantic Understanding

Future research in semantic understanding within SLAM systems is poised to address critical challenges and opportunities, enhancing robustness, adaptability, and computational efficiency. A promising direction involves estimating the full pose of semantic objects and refining data association processes for past keyframes, significantly improving the accuracy and reliability of SLAM in dynamic environments [93]. Additionally, enhancing model efficiency and robustness, as well as exploring applications in more complex environments, could lead to significant advancements in semantic understanding [82].

Integrating probabilistic data association methods into comprehensive semantic SLAM frameworks, along with exploring alternative assignment models, represents a significant research area, offering potential improvements in navigation and mapping precision [65]. Extending frameworks like STAM to support a wider range of tasks and improve human-robot interactions by better interpreting spatial semantics is another promising direction [50].

Probabilistic fusion of corrected lidar points with semantically labeled images can enhance scene understanding and support advanced navigation algorithms, which is crucial for robust performance in real-world applications [52]. Expanding the utility of proposed representations for a broader range of tasks beyond semantic segmentation could also enhance SLAM functionality and adaptability [9].

Future research directions include enhancing the robustness of Semantic Loop Closure (SLC) modules and exploring SLAM applications in various real-world scenarios, such as infrastructure inspection [45]. Integrating additional sensor configurations and expanding datasets to include varied weather conditions could provide valuable benchmarks for evaluating and refining semantic understanding techniques [42].

Improving the robustness of adaptive navigation schemes in variable conditions, such as underwater environments, aligns with potential future research directions in semantic understanding [7]. Furthermore, optimizing computational efficiency and exploring tighter integration of additional sensing modalities could enhance performance across diverse environments [59].

These research directions present substantial opportunities for enhancing semantic understanding in SLAM systems, which is critical for developing more intelligent, adaptable, and efficient autonomous robotic systems. Incorporating advanced techniques such as multi-modal semantic frameworks, object-level data association, and probabilistic data association methods enables robots to navigate complex and dynamic environments effectively. This progress not only improves localization and mapping accuracy in the presence of moving objects but also facilitates the generation of semantically enriched maps that support intelligent navigation behaviors, leading to more robust and reliable autonomous operations in real-world applications [75, 33, 38, 11, 13].

8.5 Enhancing Robotic Capabilities with Foundation Models

Foundation models represent a transformative advancement in enhancing robotic systems' capabilities, offering a unified framework for understanding and interacting with diverse environments. These models utilize advanced large-scale pre-trained architectures, particularly transformers, to achieve a nuanced understanding of visual, linguistic, and sensory data, enabling applications in scene classification, embodied vision-language planning, and robotic manipulation. By integrating multimodal machine learning techniques, these models effectively analyze and interpret complex interactions in dynamic environments, facilitating tasks requiring real-time decision-making and adaptability [114, 115, 54]. The integration of foundation models into robotic systems enables more robust perception and decision-making capabilities, allowing effective operation in complex and dynamic environments.

A key benefit of foundation models is their ability to generalize across different tasks and environments, reducing the need for extensive task-specific training. This capability is particularly valuable in scenarios where robots must adapt to new and unforeseen situations, such as emergency response or exploration in unknown terrains. By leveraging pre-trained knowledge from foundation models, robots can significantly improve their ability to recognize and interpret a wide range of sensory inputs, enabling more informed and adaptive behaviors across various tasks such as manipulation, navigation, and planning. This enhancement is facilitated by integrating advanced models like Large Language Models (LLMs) and Vision-Language Models (VLMs), which allow robots to derive actionable

AI-generated, for reference only.

insights from complex data and respond dynamically to their environments, ultimately advancing the field of general embodied Artificial Intelligence [115, 116, 61, 20, 62].

The use of foundation models in robotics also supports the integration of multi-modal data, enabling a more holistic understanding of the environment. This integration allows robots to synthesize visual, auditory, and tactile data, forming a comprehensive representation of their environment. This capability is essential for executing complex tasks such as navigating diverse settings, manipulating various objects with precision, and engaging in effective human-robot interactions, all benefiting from advanced systems like Open Knowledge-based frameworks, large language models, and dynamic visual place recognition techniques [57, 20, 115]. Processing and interpreting multi-modal data in real-time enhances the robot's situational awareness and decision-making capabilities, leading to more efficient and effective operation.

Foundation models, such as LLMs and VLMs, significantly enhance the scalability and efficiency of robotic systems by enabling flexible applications across various tasks and modalities. These models facilitate the integration of advanced perception, motion planning, and control capabilities, allowing robots to perform complex manipulation tasks based on natural language instructions. Furthermore, adopting foundation models alongside traditional learning methods can improve generalization performance and adaptability in dynamic environments, addressing critical challenges in real-world robotic applications [115, 61, 20, 117, 62]. By leveraging shared representations and pre-trained knowledge, these models can reduce the computational and data requirements for training and deploying robotic applications. This scalability is essential for deploying robotic systems across various domains, from industrial automation to healthcare and service robotics.

The integration of foundation models into robotic systems holds significant promise for advancing robotics, enabling more intelligent, adaptable, and capable autonomous systems. As advancements in research and development continue, foundation models—such as LLMs and VLMs—are set to influence the future of robotic perception and interaction significantly. These models, trained on extensive datasets, enable flexible applications across various tasks and modalities, enhancing capabilities in manipulation, navigation, and reasoning. Their integration into robotic systems can replace specific components, leading to innovative applications and improved performance in autonomous systems. For instance, the OK-Robot framework demonstrates how combining vision and language models can achieve remarkable success rates in real-world tasks, highlighting the potential for foundation models to revolutionize robotics through enhanced perception and interaction capabilities. Ongoing exploration of these technologies promises to unlock new possibilities and address challenges in the field [20, 61, 62].

9 Conclusion

This survey provides a thorough examination of Semantic SLAM and 3D Mapping, underscoring their pivotal roles in enhancing robotic autonomy. The integration of semantic information within SLAM systems significantly improves cognitive and operational capabilities, enabling robots to navigate and interact with complex environments with greater precision and understanding. Advanced algorithms and multi-sensor data fusion have been essential in addressing challenges posed by dynamic environments, thereby enhancing the robustness and efficiency of SLAM systems. Recent experiments illustrate that these integrations markedly reduce drift during UAV navigation and enhance the accuracy of semantic mapping [74].

The interdisciplinary nature of Semantic SLAM and 3D Mapping—spanning computer vision, robotics, and artificial intelligence—continues to foster innovations that broaden the functionality and adaptability of robotic systems. Ongoing advancements in AI and machine learning, particularly through deep learning techniques, are set to further augment the perceptual and navigational capabilities of autonomous systems. Such innovations present substantial opportunities for future research aimed at developing more intelligent, adaptable, and efficient robotic systems capable of operating in diverse and challenging environments.

This survey emphasizes the significance of Semantic SLAM and 3D Mapping in the evolution of intelligent robotic systems that can autonomously navigate and execute complex tasks. As the field progresses, the potential for future innovations remains extensive, promising to unveil new applications and capabilities in autonomous systems. The integration of foundation models, along with the exploration of novel sensor modalities and algorithmic strategies, will be vital in shaping the

AI-generated, for reference only.

future of robotic perception and interaction, paving the way for groundbreaking advancements in this interdisciplinary domain.

References

[1] John McCormac, Ankur Handa, Andrew Davison, and Stefan Leutenegger. Semanticfusion: Dense 3d semantic mapping with convolutional neural networks. In *2017 IEEE International Conference on Robotics and automation (ICRA)*, pages 4628–4635. IEEE, 2017.

[2] Marcela Carvalho, Maxime Ferrera, Alexandre Boulch, Julien Moras, Bertrand Le Saux, and Pauline Trouvé-Peloux. Technical report: Co-learning of geometry and semantics for online 3d mapping, 2019.

[3] Siting Zhu, Renjie Qin, Guangming Wang, Jiuming Liu, and Hesheng Wang. Semgauss-slam: Dense semantic gaussian splatting slam. *arXiv preprint arXiv:2403.07494*, 2024.

[4] Yi Zhou, Guillermo Gallego, and Shaojie Shen. Event-based stereo visual odometry, 2021.

[5] Seongwook Yoon, Jaehyun Kim, and Sanghoon Sull. Inertial guided uncertainty estimation of feature correspondence in visual-inertial odometry/slam, 2023.

[6] Xu Liu, Jiuzhou Lei, Ankit Prabhu, Yuezhan Tao, Igor Spasojevic, Pratik Chaudhari, Nikolay Atanasov, and Vijay Kumar. Slideslam: Sparse, lightweight, decentralized metric-semantic slam for multi-robot navigation, 2024.

[7] Arturo Gomez Chavez, Qingwen Xu, Christian A. Mueller, Sören Schwertfeger, and Andreas Birk. Adaptive navigation scheme for optimal deep-sea localization using multimodal perception cues, 2019.

[8] Peter Mortimer and Mirko Maehlisch. Survey on datasets for perception in unstructured outdoor environments, 2024.

[9] Md. Alimoor Reza and Jana Kosecka. Reinforcement learning for semantic segmentation in indoor scenes, 2016.

[10] Kaiqi Chen, Jialing Liu, Qinying Chen, Zhenhua Wang, and Jianhua Zhang. Accurate object association and pose updating for semantic slam. *IEEE Transactions on Intelligent Transportation Systems*, 23(12):25169–25179, 2022.

[11] Han Wang, Jing Ying Ko, and Lihua Xie. Multi-modal semantic slam for complex dynamic environments, 2022.

[12] Mirco Colosi, Irvin Aloise, Tiziano Guadagnino, Dominik Schlegel, Bartolomeo Della Corte, Kai O. Arras, and Giorgio Grisetti. Plug-and-play slam: A unified slam architecture for modularity and ease of use, 2020.

[13] Kevin J Doherty, David P Baxter, Edward Schneeweiss, and John J Leonard. Probabilistic data association via mixture models for robust semantic slam. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1098–1104. IEEE, 2020.

[14] Yiyi Liao, Jun Xie, and Andreas Geiger. Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d, 2022.

[15] Huajian Huang, Longwei Li, Hui Cheng, and Sai-Kit Yeung. Photo-slam: Real-time simultaneous localization and photorealistic mapping for monocular, stereo, and rgb-d cameras, 2024.

[16] Zirui Zhao, Yijun Mao, Yan Ding, Pengju Ren, and Nanning Zheng. Visual semantic slam with landmarks for large-scale outdoor environment, 2020.

[17] Andrei Potapov, Mikhail Kurenkov, Pavel Karpyshev, Evgeny Yudin, Alena Savinykh, Evgeny Kruzhkov, and Dzmitry Tsetserukou. Hierarchical visual localization based on sparse feature pyramid for adaptive reduction of keypoint map size, 2023.

[18] Marcus Abate, Yun Chang, Nathan Hughes, and Luca Carlone. Kimera2: Robust and accurate metric-semantic slam in the real world, 2024.

[19] Shervin Dehghani, Michael Sommersperger, Peiyao Zhang, Alejandro Martin-Gomez, Benjamin Busam, Peter Gehlbach, Nassir Navab, M. Ali Nasseri, and Iulian Iordachita. Robotic navigation autonomy for subretinal injection via intelligent real-time virtual ioct volume slicing, 2023.

[20] Peiqi Liu, Yaswanth Orru, Jay Vakil, Chris Paxton, Nur Muhammad Mahi Shafiullah, and Lerrel Pinto. Ok-robot: What really matters in integrating open-knowledge models for robotics, 2024.

[21] Eungchang Mason Lee, Junho Choi, Hyungtae Lim, and Hyun Myung. Real: Rapid exploration with active loop-closing toward large-scale 3d mapping using uavs, 2021.

[22] Peng Huang, Liekang Zeng, Xu Chen, Ke Luo, Zhi Zhou, and Shuai Yu. Edge robotics: Edge-computing-accelerated multi-robot simultaneous localization and mapping, 2022.

[23] Akira Taniguchi, Shota Isobe, Lotfi El Hafi, Yoshinobu Hagiwara, and Tadahiro Taniguchi. Autonomous planning based on spatial concepts to tidy up home environments with service robots, 2021.

[24] Juana Valeria Hurtado and Abhinav Valada. Semantic scene segmentation for robotics, 2024.

[25] Nanda Febri Istighfarin and HyungGi Jo. Leveraging spatial attention and edge context for optimized feature selection in visual localization, 2024.

[26] Daniel Gutierrez-Gomez and Jose J. Guerrero. Rgbid-slam for accurate real-time localisation and 3d mapping, 2018.

[27] Yasaman Haghighi, Suryansh Kumar, Jean-Philippe Thiran, and Luc Van Gool. Neural implicit dense semantic slam, 2023.

[28] Yuhao Zhou, Huanhuan Fan, Shuang Gao, Yuchen Yang, Xudong Zhang, Jijunnan Li, and Yandong Guo. Retrieval and localization with observation constraints, 2021.

[29] Sourav Garg, Adam Jacobson, Swagat Kumar, and Michael Milford. Improving condition- and environment-invariant place recognition with semantic place categorization, 2017.

[30] Hideki Deguchi, Kazuki Shibata, and Shun Taguchi. Language to map: Topological map generation from natural language path instructions, 2024.

[31] Vasileios Vasilopoulos, Georgios Pavlakos, Sean L. Bowman, J. Diego Caporale, Kostas Daniilidis, George J. Pappas, and Daniel E. Koditschek. Technical report: Reactive semantic planning in unexplored semantic environments using deep perceptual feedback, 2020.

[32] Jongmin Jeong, Tae Sung Yoon, and Jin Bae Park. Multimodal sensor-based semantic 3d mapping for a large-scale environment, 2018.

[33] Lukas Bernreiter, Abel Gawel, Hannes Sommer, Juan Nieto, Roland Siegwart, and Cesar Cadena. Multiple hypothesis semantic mapping for robust data association, 2020.

[34] Siting Zhu, Renjie Qin, Guangming Wang, Jiuming Liu, and Hesheng Wang. Semgauss-slam: Dense semantic gaussian splatting slam, 2024.

[35] Yulun Tian, Yun Chang, Fernando Herrera Arias, Carlos Nieto-Granda, Jonathan P. How, and Luca Carlone. Kimera-multi: Robust, distributed, dense metric-semantic slam for multi-robot systems, 2021.

[36] Guillaume Bono, Hervé Poirier, Leonid Antsfeld, Gianluca Monaci, Boris Chidlovskii, and Christian Wolf. Learning to navigate efficiently and precisely in real environments, 2024.

[37] Kevin M. Judd and Jonathan D. Gammell. The oxford multimotion dataset: Multiple se(3) motions with ground truth, 2019.

[38] a++: Efficient lidar-based sema.

[39] Zirui Zhao, Yijun Mao, Yan Ding, Pengju Ren, and Nanning Zheng. Visual-based semantic slam with landmarks for large-scale outdoor environment. In *2019 2nd China symposium on cognitive computing and hybrid intelligence (CCHI)*, pages 149–154. IEEE, 2019.

[40] Banghe Wu, Chengzhong Xu, and Hui Kong. Lidar road-atlas: An efficient map representation for general 3d urban environment, 2023.

[41] Jean-Emmanuel Deschaud. Imls-slam: Scan-to-model matching based on 3d data. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2480–2485. IEEE, 2018.

[42] Mingi Jeong, Arihant Chadda, Ziang Ren, Luyang Zhao, Haowen Liu, Monika Roznere, Aiwei Zhang, Yitao Jiang, Sabriel Achong, Samuel Lensgraf, and Alberto Quattrini Li. Seepersea: Multi-modal perception dataset of in-water objects for autonomous surface vehicles, 2025.

[43] Depanshu Sani and Saket Anand. Graph-based multi-modal sensor fusion for autonomous driving, 2024.

[44] Daniel Duberg and Patric Jensfelt. Ufomap: An efficient probabilistic 3d mapping framework that embraces the unknown, 2020.

[45] Yuezhan Tao, Xu Liu, Igor Spasojevic, Saurav Agarwal, and Vijay Kumar. 3d active metric-semantic slam. *IEEE Robotics and Automation Letters*, 9(3):2989–2996, 2024.

[46] Paul Ruvolo, Ayush Chakraborty, Rucha Dave, Richard Li, Duncan Mazza, Xierui Shen, Raiyan Siddique, and Krishna Suresh. The invisible map: Visual-inertial slam with fiducial markers for smartphone-based indoor navigation, 2023.

[47] Weikun Zhen and Sebastian Scherer. A unified 3d mapping framework using a 3d or 2d lidar, 2018.

[48] Xieyuanli Chen, Andres Milioto, Emanuele Palazzolo, Philippe Giguère, Jens Behley, and Cyrill Stachniss. Suma++: Efficient lidar-based semantic slam, 2021.

[49] Ghanta Sai Krishna, Kundrapu Supriya, and Sabur Baidya. 3ds-slam: A 3d object detection based semantic slam towards dynamic indoor environments, 2023.

[50] Francesco Riccio, Roberto Capobianco, Marc Hanheide, and Daniele Nardi. Stam: A framework for spatio-temporal affordance maps, 2016.

[51] Mathieu Gonzalez, Eric Marchand, Amine Kacete, and Jérôme Royan. Twistslam++: Fusing multiple modalities for accurate dynamic semantic slam, 2023.

[52] Mao Shan, Julie Stephany Berrio, Stewart Worrall, and Eduardo Nebot. Probabilistic egocentric motion correction of lidar point cloud and projection to camera images for moving platforms, 2020.

[53] Yiannis Kantaros, Samarth Kalluraya, Qi Jin, and George J. Pappas. Perception-based temporal logic planning in uncertain semantic maps, 2022.

[54] Zachary A Daniels and Dimitris Metaxas. A dynamic data driven approach for explainable scene understanding, 2022.

[55] Christian Häne, Lionel Heng, Gim Hee Lee, Friedrich Fraundorfer, Paul Furgale, Torsten Sattler, and Marc Pollefeys. 3d visual perception for self-driving cars using a multi-camera system: Calibration, mapping, localization, and obstacle detection, 2017.

[56] Yifan Zhu, Lingjuan Miao, Haitao Wu, Zhiqiang Zhou, Weiyi Chen, and Longwen Wu. Semantic object-level modeling for robust visual camera relocalization, 2024.

[57] Maria Waheed, Michael Milford, Klaus McDonald-Maier, and Shoaib Ehsan. Switchhit: A probabilistic, complementarity-based switching system for improved visual place recognition in changing environments, 2022.

[58] Ivan Moskalenko, Anastasiia Kornilova, and Gonzalo Ferrer. Visual place recognition for aerial imagery: A survey, 2024.

[59] M. Palieri, B. Morrell, A Thakur, K. Ebadi, J. Nash, A. Chatterjee, C. Kanellakis, L. Carlone, C. Guaragnella, and A. Agha-mohammadi. Locus: A multi-sensor lidar-centric solution for high-precision odometry and 3d mapping in real-time, 2020.

[60] Dominic Dall'Osto, Tobias Fischer, and Michael Milford. Fast and robust bio-inspired teach and repeat navigation, 2021.

[61] Xuan Xiao, Jiahang Liu, Zhipeng Wang, Yanmin Zhou, Yong Qi, Qian Cheng, Bin He, and Shuo Jiang. Robot learning in the era of foundation models: A survey, 2023.

[62] Kento Kawaharazuka, Tatsuya Matsushima, Andrew Gambardella, Jiaxian Guo, Chris Paxton, and Andy Zeng. Real-world robot applications of foundation models: A review, 2024.

[63] Tengteng Huang, Zhe Liu, Xiwu Chen, and Xiang Bai. Epnet: Enhancing point features with image semantics for 3d object detection. In *Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, part XV 16*, pages 35–52. Springer, 2020.

[64] Sara Hatami Gazani, Fardad Dadboud, Miodrag Bolic, Iraj Mantegh, and Homayoun Najjaran. Object semantics give us the depth we need: Multi-task approach to aerial depth completion, 2023.

[65] Elad Michael, Tyler Summers, Tony A. Wood, Chris Manzie, and Iman Shames. Probabilistic data association for semantic slam at scale, 2022.

[66] Tiziano Guadagnino, Benedikt Mersch, Ignacio Vizzo, Saurabh Gupta, Meher V. R. Malladi, Luca Lobefaro, Guillaume Doisy, and Cyrill Stachniss. Kinematic-icp: Enhancing lidar odometry with kinematic constraints for wheeled mobile robots moving on planar surfaces, 2025.

[67] Karel Zimmermann, Tomas Petricek, Vojtech Salansky, and Tomas Svoboda. Learning for active 3d mapping, 2017.

[68] Kurran Singh, Jungseok Hong, Nicholas R. Rypkema, and John J. Leonard. Opti-acoustic semantic slam with unknown objects in underwater environments, 2024.

[69] Fetullah Atas, Grzegorz Cielniak, and Lars Grimstad. Elevation state-space: Surfel-based navigation in uneven environments for mobile robots, 2022.

[70] Yusheng Wang, Weiwei Song, Yi Zhang, Fei Huang, Zhiyong Tu, Ruoying Li, Shimin Zhang, and Yidong Lou. Four years of multi-modal odometry and mapping on the rail vehicles, 2023.

[71] Lukas Schmid, Mansoor Nasir Cheema, Victor Reijgwart, Roland Siegwart, Federico Tombari, and Cesar Cadena. Sc-explorer: Incremental 3d scene completion for safe and efficient exploration mapping and planning, 2022.

[72] Maozhen Wang, Xianchao Long, Peng Chang, and Taskin Padir. Autonomous robot navigation with rich information mapping in nuclear storage environments, 2019.

[73] John Keller and Sebastian Scherer. A stereo algorithm for thin obstacles and reflective objects, 2019.

[74] Xu Liu, Guilherme V. Nardari, Fernando Cladera Ojeda, Yuezhan Tao, Alex Zhou, Thomas Donnelly, Chao Qu, Steven W. Chen, Roseli A. F. Romero, Camillo J. Taylor, and Vijay Kumar. Large-scale autonomous flight with real-time semantic slam under dense forest canopy, 2023.

[75] Zhentian Qian, Kartik Patath, Jie Fu, and Jing Xiao. Semantic slam with autonomous object-level data association, 2020.

[76] Kaiyu Zheng, Andrzej Pronobis, and Rajesh P. N. Rao. Learning graph-structured sum-product networks for probabilistic semantic maps, 2017.

[77] Bin Yang, Ming Liang, and Raquel Urtasun. Hdnet: Exploiting hd maps for 3d object detection. In *Conference on Robot Learning*, pages 146–155. PMLR, 2018.

[78] Yanmin Wu, Yunzhou Zhang, Delong Zhu, Yonghui Feng, Sonya Coleman, and Dermot Kerr. Eao-slam: Monocular semi-dense object slam based on ensemble data association, 2020.

[79] Kavisha Vidanapathirana, Peyman Moghadam, Ben Harwood, Muming Zhao, Sridha Sridharan, and Clinton Fookes. Locus: Lidar-based place recognition using spatiotemporal higher-order pooling, 2021.

[80] Jesús Martínez-Gómez, Vicente Morell, Miguel Cazorla, and Ismael García-Varea. Semantic localization in the pcl library, 2016.

[81] Xiao Han and Lu Yang. Sq-slam: Monocular semantic slam based on superquadric object representation, 2022.

[82] Yilun Du, Tomas Lozano-Perez, and Leslie Kaelbling. Learning object-based state estimators for household robots, 2022.

[83] Giseop Kim and Ayoung Kim. Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4802–4809. IEEE, 2018.

[84] Mickey Li, Noyan Songur, Pavel Orlov, Stefan Leutenegger, and A Aldo Faisal. Towards an embodied semantic fovea: Semantic 3d scene reconstruction from ego-centric eye-tracker videos, 2018.

[85] Hartmut Surmann, Nils Berninger, and Rainer Worst. 3d mapping for multi hybrid robot cooperation, 2019.

[86] Kenny Chen, Ryan Nemiroff, and Brett T. Lopez. Direct lidar-inertial odometry and mapping: Perceptive and connective slam, 2023.

[87] Omar Mohamed, Zeyad Mohsen, Mohamed Wageeh, and Mohamed Hegazy. Autonomous navigation in dynamic environments: Deep learning-based approach, 2021.

[88] Jiawei Hou, Yijun Yuan, and Sören Schwertfeger. Topological area graph generation and its application to path planning, 2018.

[89] Brian H. Wang, Beatriz Asfora, Rachel Zheng, Aaron Peng, Jacopo Banfi, and Mark Campbell. Multiple-hypothesis path planning with uncertain object detections, 2023.

[90] Ruiji Liu, Francisco Yandun, and George Kantor. Towards over-canopy autonomous navigation: Crop-agnostic lidar-based crop-row detection in arable fields, 2024.

[91] Scott Fredriksson, Akshit Saradagi, and George Nikolakopoulos. Voxel map to occupancy map conversion using free space projection for efficient map representation for aerial and ground robots, 2024.

[92] Chaoqun Wang, Lili Meng, Sizhen She, Ian M. Mitchell, Teng Li, Frederick Tung, Weiwei Wan, Max. Q. H. Meng, and Clarence W. de Silva. Autonomous mobile robot navigation in uneven and unstructured indoor environments, 2017.

[93] Sean L Bowman, Nikolay Atanasov, Kostas Daniilidis, and George J Pappas. Probabilistic data association for semantic slam. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 1722–1729. IEEE, 2017.

[94] Manh Do Duc, Thanh Nguyen Canh, Minh DoNgoc, and Xiem HoangVan. Fusion lidar-inertial-encoder data for high-accuracy slam, 2024.

[95] William Dubois, Matěj Boxan, Johann Laconte, and François Pomerleau. 3d mapping of glacier moulins: Challenges and lessons learned, 2024.

[96] Samuel Triest, David D. Fan, Sebastian Scherer, and Ali-Akbar Agha-Mohammadi. Unrealnet: Learning uncertainty-aware navigation features from high-fidelity scans of real environments, 2024.

[97] Cong Gao, Dinesh Rabindran, and Omid Mohareri. Rgb-d semantic slam for surgical robot navigation in the operating room, 2022.

[98] Emil Wiman, Ludvig Widén, Mattias Tiger, and Fredrik Heintz. Autonomous 3d exploration in large-scale environments with dynamic obstacles, 2023.

[99] Vandita Shukla, Luca Morelli, Pawel Trybala, Fabio Remondino, Wentian Gan, Yifei Yu, and Xin Wang. Exploring the potential of collaborative uav 3d mapping in kenyan savanna for wildlife research, 2024.

[100] Matěj Boxan, Alexander Krawciw, Effie Daum, Xinyuan Qiao, Sven Lilge, Timothy D. Barfoot, and François Pomerleau. Fomo: A proposal for a multi-season dataset for robot navigation in forêt montmorency, 2024.

[101] Kaiqi Chen, Jialing Liu, Qinying Chen, Zhenhua Wang, and Jianhua Zhang. Accurate object association and pose updating for semantic slam, 2021.

[102] Vartika Sengar, Aditya Kapoor, Nijil George, Vighnesh Vatsal, Jayavardhana Gubbi, Balamuralidhar P, and Arpan Pal. Challenges in applying robotics to retail store management, 2022.

[103] Ciro Potena, Raghav Khanna, Juan Nieto, Roland Siegwart, Daniele Nardi, and Alberto Pretto. Agricolmap: Aerial-ground collaborative 3d mapping for precision farming, 2019.

[104] Jiacheng Yuan, Jungseok Hong, Junaed Sattar, and Volkan Isler. Row-slam: Under-canopy cornfield semantic slam, 2021.

[105] Joao Marcos Correia Marques, Albert Zhai, Shenlong Wang, and Kris Hauser. On the overconfidence problem in semantic 3d mapping, 2023.

[106] Weihan Wang, Bharat Joshi, Nathaniel Burgdorfer, Konstantinos Batsos, Alberto Quattrini Li, Philippos Mordohai, and Ioannis Rekleitis. Real-time dense 3d mapping of underwater environments, 2023.

[107] Callum Rhodes, Cunjia Liu, and Wen-Hua Chen. Structurally aware 3d gas distribution mapping using belief propagation: a real-time algorithm for robotic deployment, 2022.

[108] Jiarong Lin and Fu Zhang. Loam livox: A fast, robust, high-precision lidar odometry and mapping package for lidars of small fov. In *2020 IEEE international conference on robotics and automation (ICRA)*, pages 3126–3131. IEEE, 2020.

[109] Kejie Qiu, Shenzhou Chen, Jiahui Zhang, Rui Huang, Le Cui, Siyu Zhu, and Ping Tan. Compact 3d map-based monocular localization using semantic edge alignment, 2021.

[110] Shan Wang, Yanhao Zhang, Ankit Vora, Akhil Perincherry, and Hongdong Li. Satellite image based cross-view localization for autonomous vehicle, 2023.

[111] Kunyi Li, Michael Niemeyer, Nassir Navab, and Federico Tombari. Dns slam: Dense neural semantic-informed slam, 2023.

[112] Luigi Nardi, Bruno Bodin, M. Zeeshan Zia, John Mawer, Andy Nisbet, Paul H. J. Kelly, Andrew J. Davison, Mikel Luján, Michael F. P. O'Boyle, Graham Riley, Nigel Topham, and Steve Furber. Introducing slambench, a performance and accuracy benchmarking methodology for slam, 2015.

[113] Xuanpeng Li and Rachid Belaroussi. Semi-dense 3d semantic mapping from monocular slam, 2016.

[114] Jonathan Francis, Nariaki Kitamura, Felix Labelle, Xiaopeng Lu, Ingrid Navarro, and Jean Oh. Core challenges in embodied vision-language planning, 2023.

[115] Wenlong Huang, Chen Wang, Ruohan Zhang, Yunzhu Li, Jiajun Wu, and Li Fei-Fei. Voxposer: Composable 3d value maps for robotic manipulation with language models. *arXiv preprint arXiv:2307.05973*, 2023.

[116] Min Zhao and Andre G. Marquez. Understanding humans' strategies in maze solving, 2013.

[117] Cyrille Berger, Patrick Doherty, Piotr Rudol, and Mariusz Wzorek. Hastily formed knowledge networks and distributed situation awareness for collaborative robotics, 2021.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.