

ABSTRACT

This research is motivated by the potential of autonomous mobile robots (AMRs) in enhancing safety, productivity, and efficiency in the construction industry. The dynamic and complex nature of construction sites presents significant challenges to AMRs, particularly in localization and mapping – a process where AMRs determine their own position in the environment while creating a map of the surrounding area. These capabilities are crucial for autonomous navigation and task execution but are inadequately addressed by existing solutions, which primarily rely on visual Simultaneous Localization and Mapping (SLAM) methods. These methods are often ineffective in construction sites due to their underlying assumption of a static environment, leading to unreliable outcomes. Therefore, there is a pressing need to enhance the applicability of AMRs in construction by addressing the limitations of current localization and mapping methods in addressing the dynamic nature of construction sites, thereby empowering AMRs to function more effectively and fully realize their potential in the construction industry.

The overarching goal of this research is to fulfill this critical need by developing a novel visual SLAM framework that is capable of not only detecting and segmenting diverse dynamic objects in construction environments but also effectively interpreting the semantic structure of the environment. Furthermore, it can efficiently integrate these functionalities into a unified system to provide an improved SLAM solution for dynamic, complex, and unstructured environments. The rationale is that such a SLAM system could effectively address the dynamic nature of construction sites, thereby significantly improving the efficiency and accuracy of robot localization and mapping in the construction working environment.

Towards this goal, three specific objectives have been formulated. The first objective is to develop a novel methodology for comprehensive dynamic object segmentation that can support visual SLAM within highly variable construction environments. This novel method integrates class-agnostic objectness masks and motion cues into video object segmentation, thereby significantly improving the identification and segmentation of dynamic objects within construction sites. These dynamic objects present a significant challenge to the reliable operation of AMRs and, by accurately identifying and segmenting them, the accuracy and reliability of SLAM-based localization is expected to greatly improve. The key to this innovative approach involves a four-stage method for dynamic object segmentation, including objectness mask generation, motion

saliency estimation, fusion of objectness masks and motion saliency, and bi-directional propagation of the fused mask. Experimental results show that the proposed method achieves a highest of 6.4% improvement for dynamic object segmentation than state-of-the-art methods, as well as lowest localization errors when integrated into visual SLAM system over public dataset.

The second objective focuses on developing a flexible, cost-effective method for semantic segmentation of construction images of structural elements. This method harnesses the power of image-level labels and Building Information Modeling (BIM) object data to replace the traditional and often labor-intensive pixel-level annotations. The hypothesis for this objective is that by fusing image-level labels with BIM-derived object information, a segmentation that is competitive with pixel-level annotations while drastically reducing the associated cost and labor intensity can be achieved. The research method involves initializing object location, extracting object information, and incorporating location priors. Extensive experiments indicate the proposed method with simple image-level labels achieves competitive results with the full pixel-level supervisions, but completely remove the need for laborious and expensive pixel-level annotations when adapting networks to unseen environments.

The third objective aims to create an efficient integration of dynamic object segmentation and semantic interpretation within a unified visual SLAM framework. It is proposed that a more efficient dynamic object segmentation with adaptively selected frames combined with the leveraging of a semantic floorplan from an as-built BIM would speed up the removal of dynamic objects and enhance localization while reducing the frequency of scene segmentation. The technical approach to achieving this objective is through two major modifications to the classic visual SLAM system: adaptive dynamic object segmentation, and semantic-based map point reliability update. Upon the accomplishment of this objective, an efficient framework is developed that seamlessly integrates dynamic object segmentation and semantic interpretation into a visual SLAM framework. Experiments demonstrate the proposed framework achieves competitive performance over the testing scenarios, with processing time almost halved than the counterpart dynamic SLAM algorithms.

In conclusion, this research contributes significantly to the adoption of AMRs in construction by tailoring a visual SLAM framework specifically for dynamic construction sites. Through the integration of dynamic object segmentation and semantic interpretation, it enhances localization accuracy, mapping efficiency, and overall SLAM performance. With broader implications of

visual SLAM algorithms such as site inspection in dangerous zones, progress monitoring, and material transportation, the study promises to advance AMR capabilities, marking a significant step towards a new era in construction automation.