Robust Real-Time 3D Mapping for Augmented Reality in Dynamic Environments

Background

Simultaneous Localization and Mapping (SLAM) is a core technology for Augmented Reality (AR). It allows a device, like a mobile phone or headset, to build a 3D map of its surroundings and simultaneously track its precise position and orientation (6-DOF pose) within that map. This capability is essential for AR applications to "anchor" virtual content to the real world, making it appear stable as the user moves. However, most traditional visual SLAM systems (e.g., ORB-SLAM2) are built on the "static world assumption," which presumes the entire environment is rigid and stationary.

Problem Specification

The static world assumption is frequently violated in real-world AR scenarios, such as homes, offices, or public spaces, which are inherently dynamic. Objects like walking people, moving vehicles, or opening doors introduce non-static elements into the scene. When a traditional SLAM algorithm uses features from these moving objects for tracking, it introduces severe errors into its pose estimation, leading to "drift." Furthermore, these dynamic elements corrupt the 3D map by creating "ghosting" artifacts, making the map unusable for long-term AR applications. This thesis will address this problem by designing a SLAM system that is robust to such dynamic content.

Suggested Method

The proposed method is to build a dynamic-robust SLAM system by augmenting a modern baseline (e.g., ORB-SLAM3). This will be achieved by integrating a real-time deep learning module to distinguish static and dynamic content.

Dynamic Feature Removal: A parallel thread will run a real-time semantic segmentation network (like Mask R-CNN) or an object detector (like YOLOv8) to identify potentially dynamic objects (e.g., "person," "car") in each frame.

Robust Tracking: Features detected on these dynamic objects will be masked and excluded from the tracking and mapping processes, ensuring that the system's pose estimation relies only on stable, static background features.

Static Map Inpainting (Extension): As an extension, the areas occluded by the removed dynamic objects will be "inpainted" using a generative model. This restores the static background, allowing the system to extract more stable features from the newly visible area and preventing tracking loss when large dynamic objects occlude the scene. The system's performance will be evaluated against the baseline on public dynamic SLAM datasets (e.g., TUM RGB-D).

Relevant Articles

[1] Bescos, B., Facil, J. M., Civera, J., & Neira, J. (2018). "DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes". IEEE Robotics and Automation Letters, 3(4), 4073-4080.

[2] Wang, Y., et al. (2024). "YDM-SLAM: YOLOv8-Powered Dynamic Mapping of Environment Using ORB-SLAM3". In 2024 4th International Conference on Data Science and Network Security (ICDSNS).

[3] Li, S., et al. (2024). "An Inpainting SLAM Approach for Detecting and Recovering Regions with Dynamic Objects". In 2024 IEEE International Conference on Robotics and Automation (ICRA).

[4] Saputra, M. R. U., et al. (2023). "A survey on real-time 3D scene reconstruction with SLAM methods in embedded systems". Journal of Real-Time Image Processing, 20(2), 26.

[5] Glick, E., et al. (2025). "SLAM- X: Generalizable Dynamic Removal for NeRF and Gaussian Splatting SLAM". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).