

Towards Online Open-vocabulary 3D Scene Graphs

Frank Gallagher¹, Marco Cagnetti² and John B. McDonald¹

¹Maynooth University, Ireland

²LAAS-CNRS/Université de Toulouse, France



Maynooth
University
National University
of Ireland Maynooth



Taighde Éireann
Research Ireland

Overview

TL;DR: Integration of ORB SLAM 3 with ConceptGraphs to achieve online open-vocabulary object mapping

In order for a robot to operate in unknown real-world environments, it must be able to reason about its surroundings at a semantic level.

Recent advances in Spatial AI have yielded systems, such as ConceptGraphs [1], that produce object-centric open-vocabulary semantic maps.

A key limitation of ConceptGraphs is its reliance on static, precomputed camera poses. This means an environment must be mapped before building the semantic representation.

This work aims to address this problem by integrating a state-of-the-art visual SLAM system (ORB-SLAM3) [2] with ConceptGraphs, enabling a robot to build a semantic map online as it explores an environment for the first time, while also updating previous sections of the map as optimised poses are made available by the underlying SLAM system.

We demonstrate this system on real-world data captured by a handheld Intel RealSense D435i camera and release the code on GitHub (scan the QR code below).

Future Work

- Extend the integration of ORB-SLAM3 to leverage its full feature set in ConceptGraphs (i.e. spawning submaps in response to tracking loss, multimap support)
- Optimise codebase to approach real-time operation
- Port codebase to ROS architecture to facilitate robot deployment
- Extend the representation to hierarchical graph structure (i.e. object-object relationships, rooms, navigability layer)

GitHub



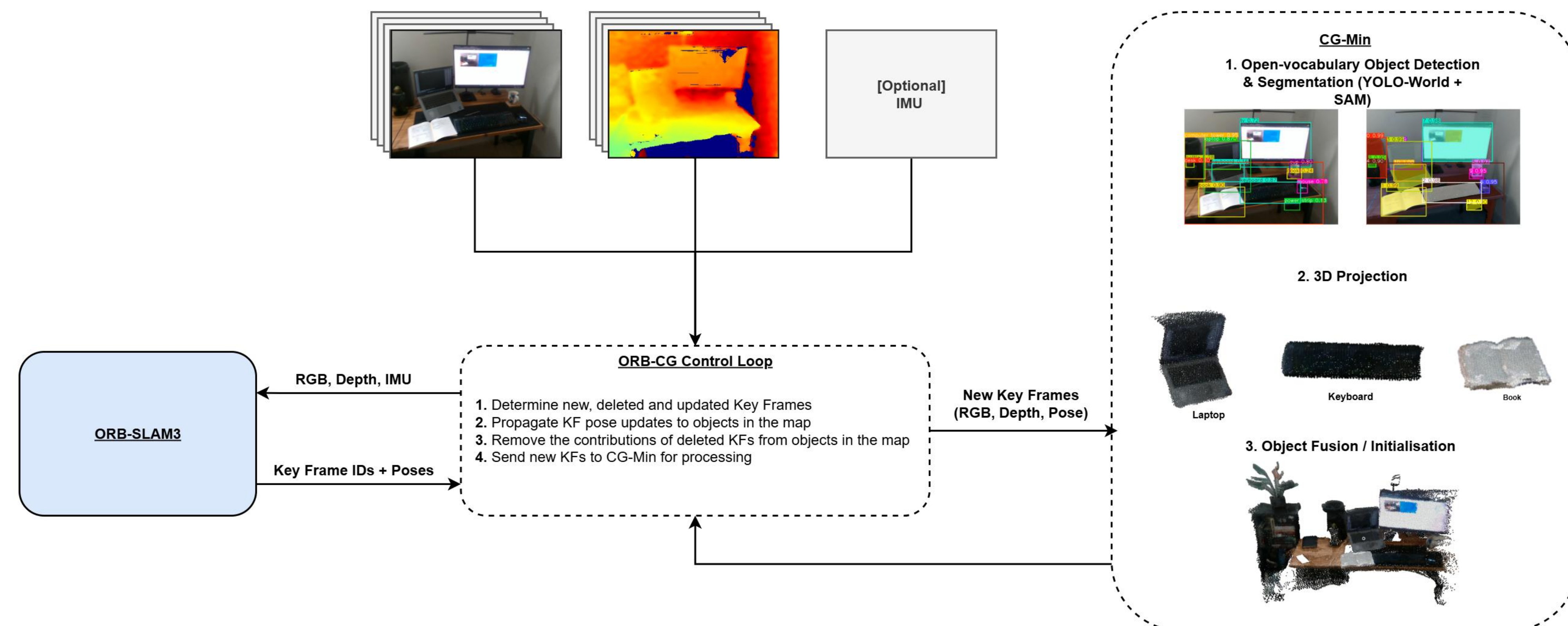
References:

[1] Gu et al., "ConceptGraphs: Open-vocabulary 3D Scene Graphs for Perception and Planning", ICRA, 2024

[2] Campos et al., "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM", IEEE Trans. Robot., 2021

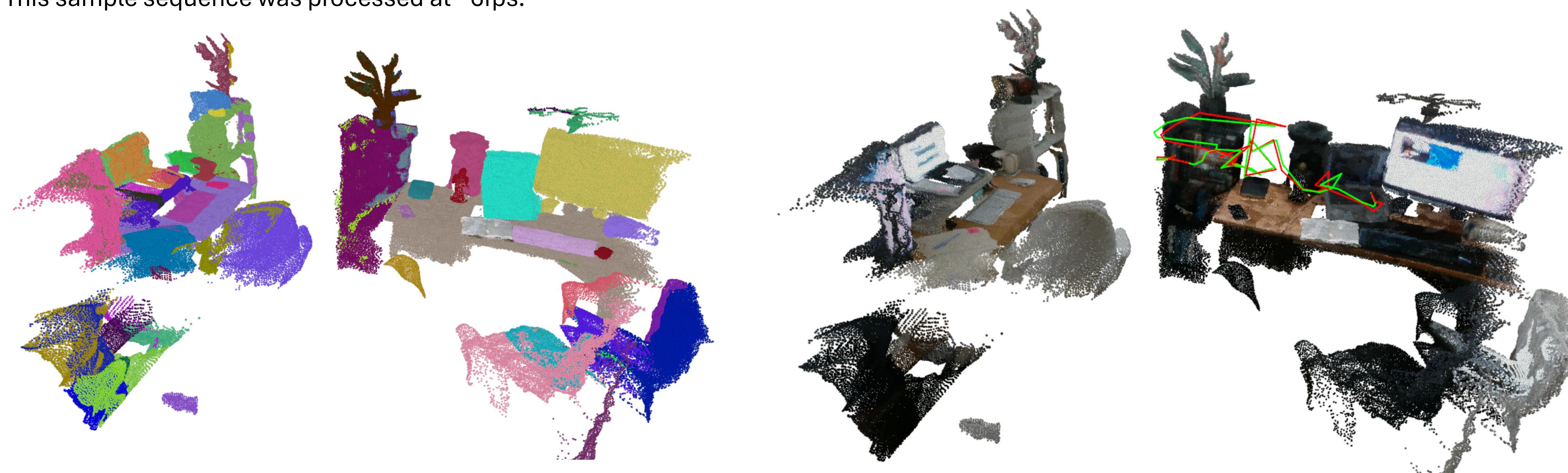
System

System consists of custom ORB-SLAM3 python bindings, custom stripped-down version of ConceptGraphs (CG-Min) and a control loop to the processing of Key Frames proposed by ORB-SLAM3.



Results

System combines the semantic mapping functionality provided by ConceptGraphs with the online camera tracking provided by ORB-SLAM3, allowing a scene to be mapped without a priori knowledge of the camera trajectory. This sample sequence was processed at ~6fps.



Instance segmentation Point Cloud

Live Camera Trajectory vs Optimised Camera Trajectory