Multi-Resolution 3D Scene Graph Construction and Optimization

Improving 3D Object Representations in the REACT Framework

Riccardo Zappa

Aalto Intelligent Robotics Group

August 22, 2025

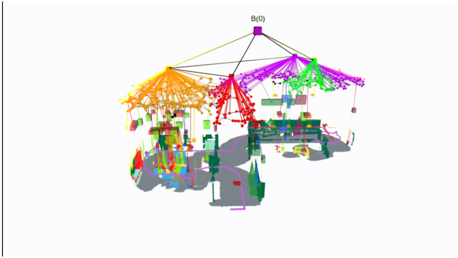
Aalto University
School of Electrical
Engineering

The Goal: From Monolithic Maps to Per Object Resolution

For a robot to be truly useful in our world, it needs more than just a map. It needs to "understand" the scene.

Project Context:

- **Hydra (MIT):** A powerful framework for large-scale 3D reconstruction. Its primary output is a 3D scene graph wiht a single, detailed 3D mesh of the environment.
- **REACT (Aalto):** Builds on Hydra to add an efficient clustering of object nodes.

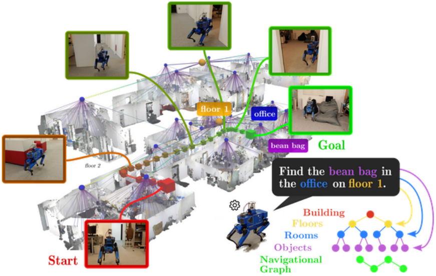


Aalto University
School of Electrical
Engineering

The Scene Graph: A Robot's Brain Map

A scene graph organizes a 3D scene into a meaningful hierarchy.

- It's not just geometry; it's a network of **nodes** (things) and **edges** (relationships).
- **Nodes**: Rooms, objects (a chair), the robot itself.
- **Edges**: Relationships like "is inside," "is on top of."



Aalto University
School of Electrical Engineering

How Hydra Currently Builds the Scene

The current pipeline relies on a monolithic (single-piece) representation.
The Process:

1. Fuse sensor data into a **Truncated Signed Distance Function (TSDF)**. Each voxel's distance value $D(x)$ is updated via a running weighted average:

$$D_{new}(x) = \frac{W_{old}(x) D_{old}(x) + w_{new} d_{new}(x)}{W_{old}(x) + w_{new}}$$

2. Extract a single, unified **3D mesh** from this volume.
3. Identify objects by segmenting or "cutting out" pieces of this global mesh.

Disadvantages of the Monolithic Approach

This "one-size-fits-all" model creates significant limitations for robotic interaction.

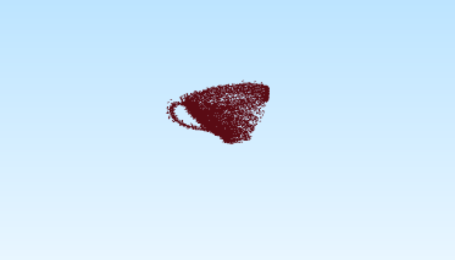
- **Problem 1: Fixed Resolution**
  - A large wall and a small, intricate coffee mug are represented with the same level of detail.
  - This is inefficient and lacks the fidelity needed for tasks like grasping.
- **Problem 2: Geometric Entanglement**
  - Objects are fundamentally "stitched into" the fabric of the world mesh.
  - Treating an object as an independent entity for analysis or manipulation is computationally expensive and clumsy.

The Core Idea: Independent Object Representations

My work decouples the object's representation from the global map.

The Goal: Instead of a coarse mesh cutout, each object node in the scene graph will now store its own **dedicated, high-fidelity point cloud**. This

allows us to maintain a lightweight global map for navigation while having rich, detailed models of objects for interaction.



Aalto University
School of Electrical
Engineering

The Theoretical Workflow

I've integrated a new pipeline for creating and refining these object-centric models.

1. **Instance Segmentation:** Identify object instances in the 2D camera image.
2. **High-Res Cloud Generation:** Project the 2D mask into 3D using depth data.
3. **Data Association (ICP):** Precisely align the new cloud with the existing model using the Iterative Closest Point algorithm. The objective is to find the rotation $\mathbf{R}$ and translation $\mathbf{t}$ that minimize the error:

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} ||(\mathbf{R}\mathbf{p}_i + \mathbf{t}) - \mathbf{q}_i||^2$$

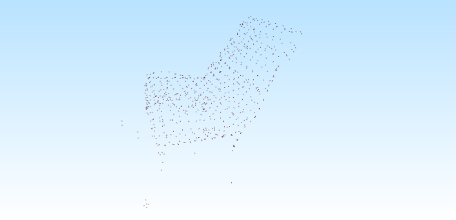
4. **Model Fusion & Refinement:** Merge the aligned cloud and apply voxel grid downsampling to maintain a consistent density.
5. **Attribute Update:** Recalculate a tighter bounding box and update the object's position.

Aalto University
School of Electrical
Engineering

Example: The Impact of High-Fidelity Models

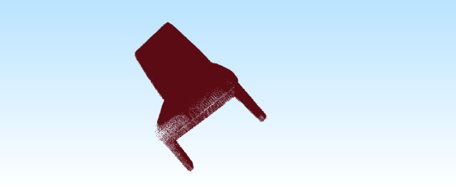
Before: Coarse Mesh Segment

- Low detail, noisy.
- Derived from global TSDF.
- Difficult for grasp planning.



After: Fused Point Cloud

- High detail, clean.
- Fused from multiple sensor views.
- Ideal for manipulation and analysis.



Aalto University
School of Electrical Engineering

Next Steps: Next Week

What could be improved?

- The ICP and downsampling parameters are currently fixed. They could be adapted based on object category (e.g., 'sofa' needs different settings than a 'cup').

Immediate Implementation Goals:

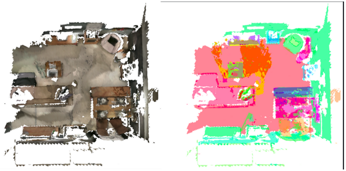
- Implement per-label parameters for ICP and downsampling to improve registration quality.
- Benchmark the performance gain in terms of memory usage and object localization accuracy.
- Implement a publisher node to visualize the incremental fusion of objects point clouds.

Aalto University
School of Electrical
Engineering

The Next Frontier: Per-Object TSDF Fusion

Per object point cloud implementation is a great achievement but for a truly solid, watertight 3D model, the next step is to give each object its own miniature **TSDF volume**.

- This allows for robust fusion of observations over time, filling in holes and removing noise.
- We can then extract a high-quality, continuous mesh for each object on demand.
- This approach is inspired by seminal works like **Panoptic Mapping (ETH Zurich)**, which demonstrates the power of this hybrid mapping strategy.



Aalto University School of Electrical Engineering

Results obtained

This shift from a monolithic to a hybrid, object-centric model is critical for the project.

- **Computational Efficiency:** Use lightweight models for navigation and high-detail models only when needed for interaction.
- **Enhanced Capability:** Enables advanced manipulation. A robot can't plan a precise grasp on a blurry model; it needs the high fidelity this method provides.
- **Scalability & Robustness:** The system can map larger, more complex environments without getting bogged down by unnecessary detail, and object models become more robust over time.

References

Key Inspirations and Frameworks:

- Rosinol, A., et al. (2022). *Hydra: A Real-time Spatial Perception System for 3D Scene Graph Construction and Optimization*. IEEE Robotics and Automation Letters. (MIT-SPARK)
- Aalto Intelligent Robotics. *REACT Project*. `github.com/aalto-intelligent-robotics/REACT`
- Schmid, L., et al. (2022). *Panoptic Multi-TSDFs: a Flexible Representation for Online Multi-resolution Volumetric Mapping...* In 2022 IEEE ICRA.

Aalto University
School of Electrical
Engineering