Real-time 3D Semantic Scene Graph Generation for Robot Manipulation

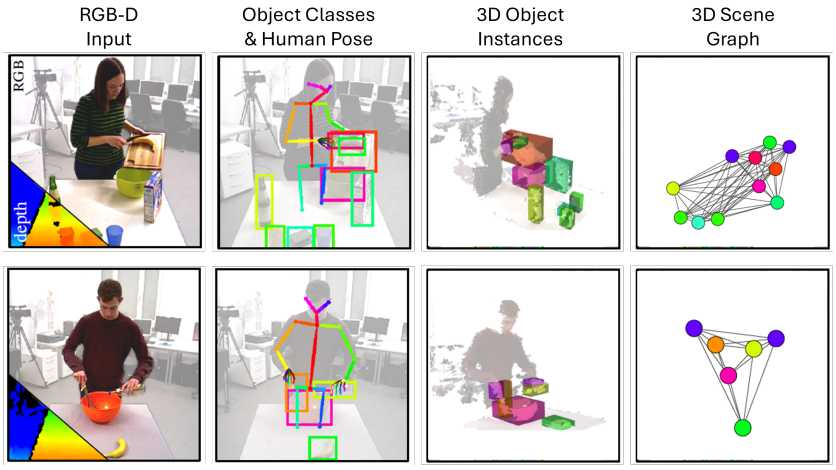

Figure 1: Two sample scenes from [1], depicting cutting and stirring manipulations performed by different subjects. From the raw RGB-D sensor input (first column), all relevant object and human pose information are extracted (second column). Subsequently, the pose of each object instance is computed (third column) to construct the final 3D scene graphs (last column).

Transforming complex 3D visual scenes into symbolic representations is a fundamental step toward grounded robotic reasoning. 3D scene graphs bridge the gap between raw sensory data and high-level understanding by capturing spatio-temporal semantic information of the observed scene. This structural abstraction is vital for robots to interpret and replicate complex human-demonstrated manipulations, such as cutting and stirring, as shown in Fig. 1.

This project focuses on developing a real-time 3D semantic scene graph generation pipeline tailored for tabletop manipulation, such as in [2, 1]. Processing RGB-D streams around 30 Hz, the pipeline will continuously map object identities and human hand poses as nodes, while edges will encode semantic embeddings of spatial and temporal relations, such as *touching, being above, moving together* and *getting close*, among others. The pipeline will integrate off-the-shelf, state-of-the-art semantic segmentation (SAM3, FastSAM), object detection (YOLO), and human pose estimation (AlphaPose) models into a ROS2-based robot architecture.

The candidate should have a solid theoretical background in deep neural networks and hands-on experience with PyTorch and ROS2.

To apply or learn more, please contact Eren Aksoy: eren.aksoy@cs.lth.se.

References

[1] C. R. G. Dreher, M. Wächter, and T. Asfour, "Learning object-action relations from bimanual human demonstration using graph networks," *IEEE Robotics and Automation Letters*, 2019.

[2] E. Erdogan, S. Sariel, and E. E. Aksoy, "Real-time manipulation action recognition with a factorized graph sequence encoder," in *IEEE/RSJ IROS*, 2025.