Multi-Modal Sensor Fusion for 3D Scene Graph Construction

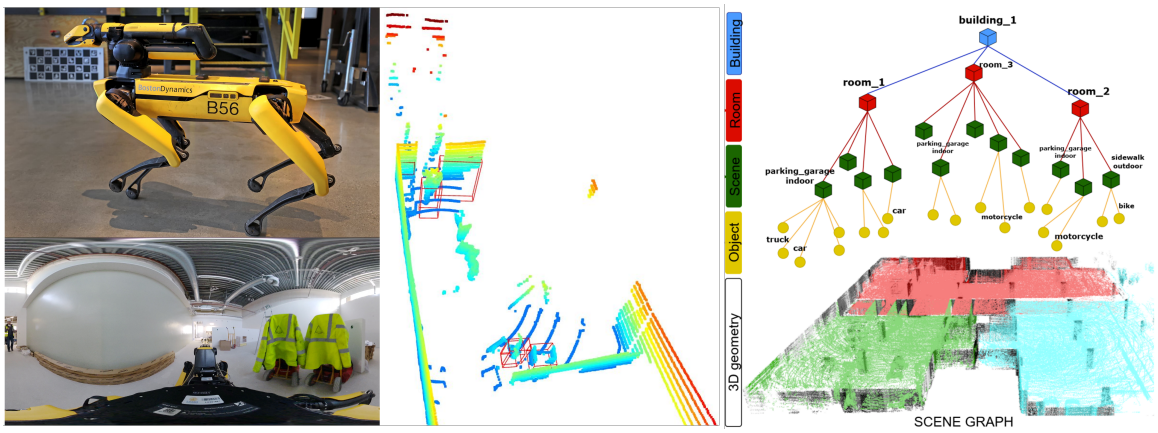

Figure 1: Spot robot platform (top-left) with its primary perception inputs: RGB camera image (bottom-left) and corresponding 3D LiDAR point cloud data with predicted object bounding boxes (middle). Fused sensor data is used to construct 3D scene graphs (right).

Autonomous mobile robots like the Boston Dynamics Spot rely on precise environmental understanding to navigate complex, unstructured terrains. While individual sensors have inherent limitations (e.g., 3D LiDAR lacks dense color information and RGB cameras struggle with direct depth precision), fusion of different sensor modalities offers more accurate and robust downstream perception solutions such as 3D object detection and semantic segmentation.

As illustrated in Fig. 1, this project focuses on studying various deep neural network-based sensor-fusion strategies, including early-, mid-, and late-fusion, for integrating LiDAR and RGB camera data to enhance 3D object detection and semantic segmentation on the Boston Dynamics Spot platform. All detected objects and segments will be further employed to construct 3D scene graphs [1, 2] (see Fig. 1-right).

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

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

References

[1] N. Hughes, Y. Chang, and L. Carlone, "Hydra: A real-time spatial perception system for 3D scene graph construction and optimization," in *Robotics: Science and Systems (RSS)*, 2022.

[2] A. Longo, C. Chung, M. Palieri, S.-K. Kim, A. Agha, C. Guaragnella, and S. Khattak, "Pixels-to-graph: Real-time integration of building information models and scene graphs for semantic-geometric human-robot understanding," in *2025 IEEE 21st International Conference on Automation Science and Engineering (CASE)*, pp. 2774–2781, 2025.