EECE 5550: Mobile Robotics
Northeastern University, Fall 2025

Project Proposal

Due Date: November 09, 2025

Raphael Dias
Ahmad Hassan
Mir Munavvar Ali

What are we proposing to do?

We propose to conduct a rigorous empirical evaluation of **DG-SLAM: Robust Dynamic Gaussian Splatting SLAM with Hybrid Pose Optimization** (Xu et al., NeurIPS 2024). This recent paper claims to be the *first* robust dynamic Gaussian splatting SLAM system capable of real-time rendering and high-fidelity reconstruction in dynamic environments. We selected this paper because it addresses a critical challenge in mobile robotics; SLAM in dynamic environments while making specific, testable performance claims about achieving state-of-the-art results. The integration of Gaussian Splatting with dynamic SLAM is particularly novel and highly relevant to current research trends. Our evaluation will focus on three key aspects:

1. Verification of whether we can reproduce the reported results on the original datasets.
2. Stress testing to examine how the system performs under conditions not tested in the paper.
3. Analysis to identify scenarios where the claimed robustness fails.

Problem Statement

Given a sequence of RGB-D frames $\{I_i, D_i\}_{i=1}^{N}$ where $I_i \in \mathbb{R}^3$ and $D_i \in \mathbb{R}$, the goal is to simultaneously:

1. Estimate camera poses: $\{\xi_i\}_{i=1}^{N}$ where $\xi_i \in SE(3)$
2. Reconstruct static 3D scene: $G = \{G_i: (\mu_i, \Sigma_i, \alpha_i, h_i)\}$ where $\mu_i \in \mathbb{R}^3, \Sigma_i \in \mathbb{R}^{3x3}, \alpha_i \in \mathbb{R}, h_i \in \mathbb{R}^{16}$

Key Algorithm Components to Evaluate

Motion Mask Generation: The paper proposes combining depth warp masks with semantic segmentation:

1. Depth warp (Eq. 5): $p_{i\to j} = KT_{ji}(K^{-1}D_i(p_i)p_i^{\text{homo}})$
2. Depth warp mask (Eq. 6): $\widehat{M}_{j,i}^{wd} : \{\cap_{p_i \in D_i} 1\left(D_j\left(p_{i\to j}\right) - D_i(p_i) < \varepsilon_{\text{th}}\right) \otimes I_{m\times n}\}$
3. Final mask (Eq. 7): $\widehat{M}_j = \widehat{M}_{j,i}^{wd} \cap \widehat{M}_{j,i-1}^{wd} \cap \ldots \cap \widehat{M}_{j,i-N}^{wd} \cup \widehat{M}_j^{sg}$

Hybrid Pose Optimization: Two-stage approach:

1. Coarse stage: DROID-SLAM visual odometry

$$E(G,d) = \Sigma_{(i,j)\in\varepsilon} \|p_{ij} * - \Pi_C(G_{ij} \circ \Pi_C^{-1}(p_i, d_i))\|^2{}_{\Sigma_{ij}\cdot M_j^c}$$

2. Fine stage: Photometric alignment with Gaussian splatting

$$\xi_i^* = \arg\min_{\xi_i} \lambda_1 \frac{1}{M} \sum_{i=1}^{M} \left\| \left(\hat{C}(\mathcal{G}, \xi_i) - C\right) \cdot \widehat{\mathcal{M}}_i \cdot \widehat{\mathcal{O}}_i \right\|_2^2 + \lambda_2 \frac{1}{N_d} \sum_{\mathbf{p}\in N_d} \left\| \left(\hat{D}(\mathcal{G}, \xi_i) - D\right) \cdot \widehat{\mathcal{M}}_i \cdot \widehat{\mathcal{O}}_i \right\|_2^2.$$

Specific Experiments

Reproduction Study: Our first goal is to verify the claimed performance on TUM RGB-D and BONN datasets. We will measure Absolute Trajectory Error (ATE) in cm, reconstruction accuracy, completion, and completion ratio, as well as runtime performance in FPS. We consider the reproduction successful if our results fall within 15% of the reported values.

Robustness Testing: We will test three scenarios not adequately covered in the paper. First, we'll examine motion blur and fast camera motion by simulating rapid movements through frame skipping at 2x and 3x speeds and adding synthetic motion blur to test images. Our hypothesis is that the system may fail when optical flow becomes unreliable. Second, we'll test semantic segmentation failure cases by running the system with degraded/noisy semantic masks (adding 20% and 40% noise) and with semantic priors disabled entirely. We hypothesize that over-reliance on semantic priors will cause tracking failures. Third, we'll test varying dynamic object densities by creating synthetic scenes with 25%, 50%, and 75% dynamic content, hypothesizing that performance degrades non-linearly with increased dynamic content.

Proposed Extensions: Based on our findings, we will implement one modest improvement. Option A involves adaptive threshold selection for the depth warp mask (currently fixed at 0.6) by implementing dynamic thresholds based on scene statistics. Option B involves confidence weighting for semantic masks, where we weigh semantic mask contributions based on segmentation confidence scores.

Expected Output

A successful execution will produce a reproducible codebase that accepts RGB-D sequences and outputs camera trajectory estimates (SE(3) poses), 3D Gaussian map representations, and rendered novel views. Our evaluation report will include reproduction results on original benchmarks, performance analysis under stress conditions, identified failure modes and limitations, and our implemented improvement with comparative results. The quantitative analysis will show performance degradation curves under various conditions, statistical significance testing of our findings, and clear documentation of scenarios where the method fails.

REFERENCES

[1] Y. Xu, H. Jiang, Z. Xiao, J. Feng, and L. Zhang, “DG-slam: Robust dynamic gaussian splatting slam with hybrid pose optimization,” arXiv.org, https://arxiv.org/abs/2411.08373 (accessed Nov. 9, 2025).