Real-Time Visual SLAM for Dynamic Environments using Hybrid Segmentation and Optical Flow

Yung-Ching Sun, Peng-Chen Chen, Chi Zhang, Hao Yin
University of Michigan, Ann Arbor, MI 48109

Introduction

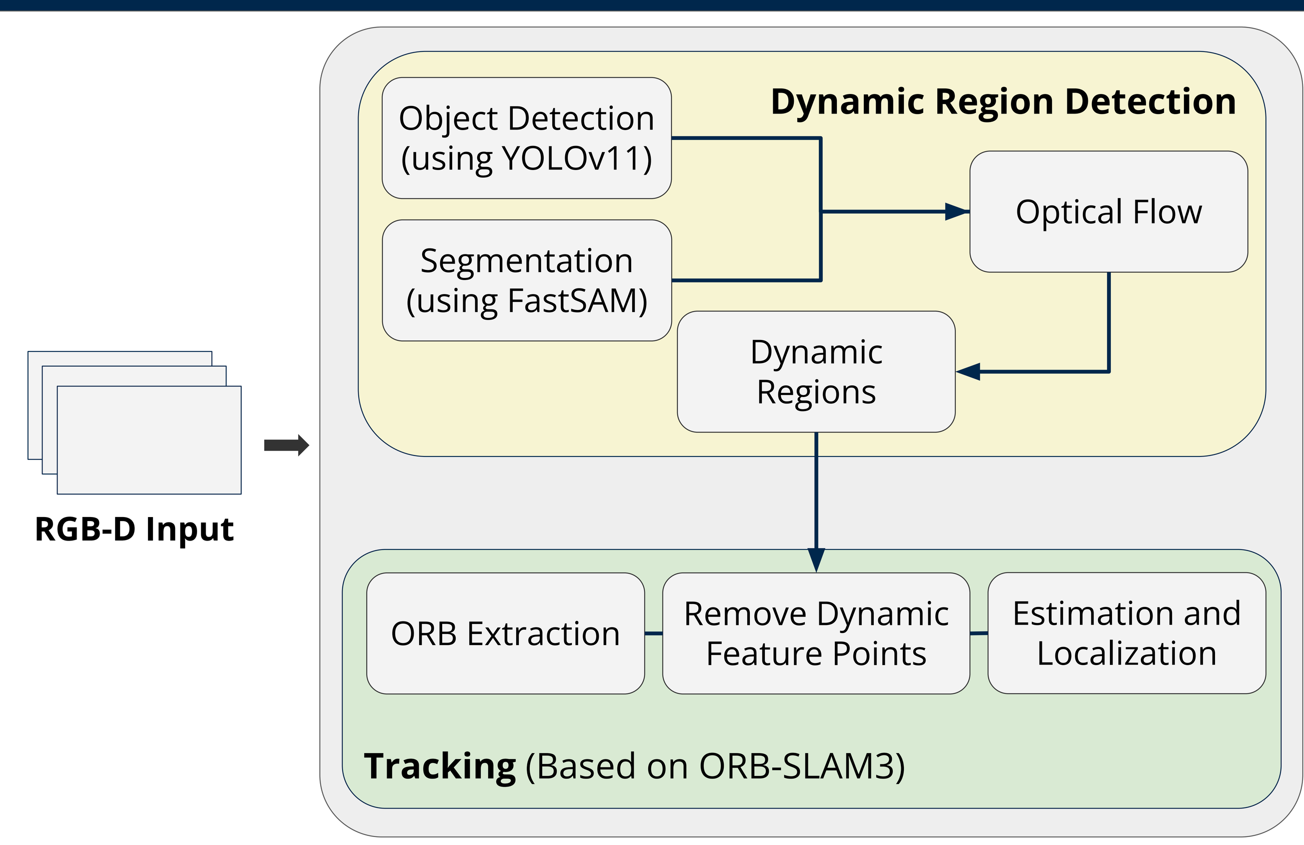
Motivation and Objectives

- Most SLAM systems are designed under the static-world assumption and perform worse in dynamic environments.
- Existing Dynamic SLAM methods focus on addressing labeled dynamic objects.
- Develop a modular, real-time pipeline that incorporate segmentation and motion analysis to mask dynamic regions to enhance ORB-SLAM3 [1] in dynamic scenes.

Novelty

- Combine **FastSAM [2]** and **YOLO11n-seg [3]** to segment both labeled and unlabeled potential dynamic regions.
- Incorporate **optical flow estimation** for precise identification of truly dynamic areas in the scene.

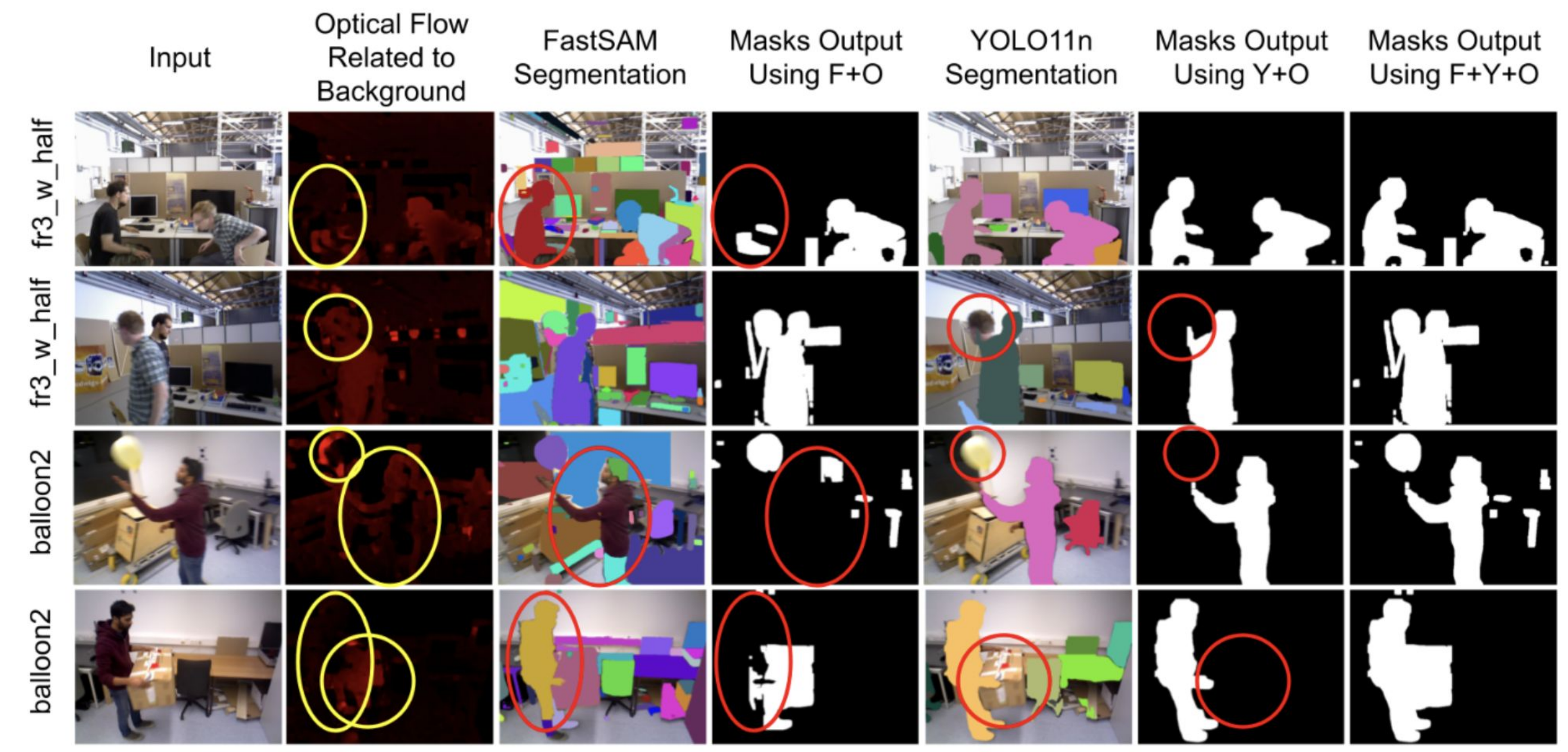
Methodology



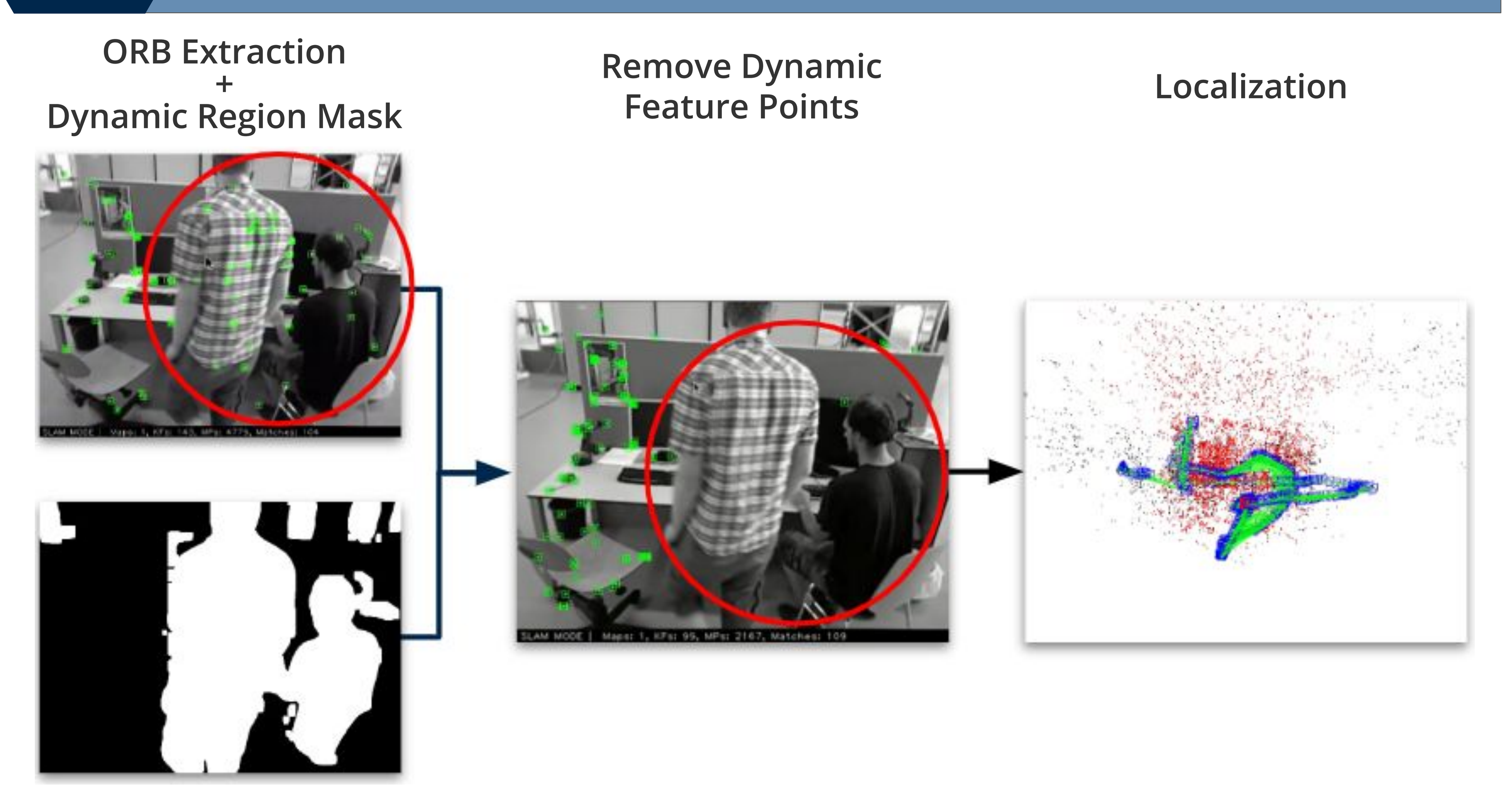
1 Dynamic Region Detection

- **Segmentation:** YOLO11n-seg [3] segments labeled dynamic classes while FastSAM [2] provides label-free segmentation.
- **Optical Flow:** Pixel-wise dense optical is computed between consecutive frames using the Farneback method.
- **Dynamic Regions Classification:** Segmented regions with higher flow magnitude than the background are classified as dynamic using adaptive thresholds: $\epsilon_{yolo} = 1.15, \epsilon_{fastsam} = 1.7$
- **Masks Generation:** Dilated masks and passed to ORB-SLAM3.

1 Dynamic Region Detection



2 Tracking

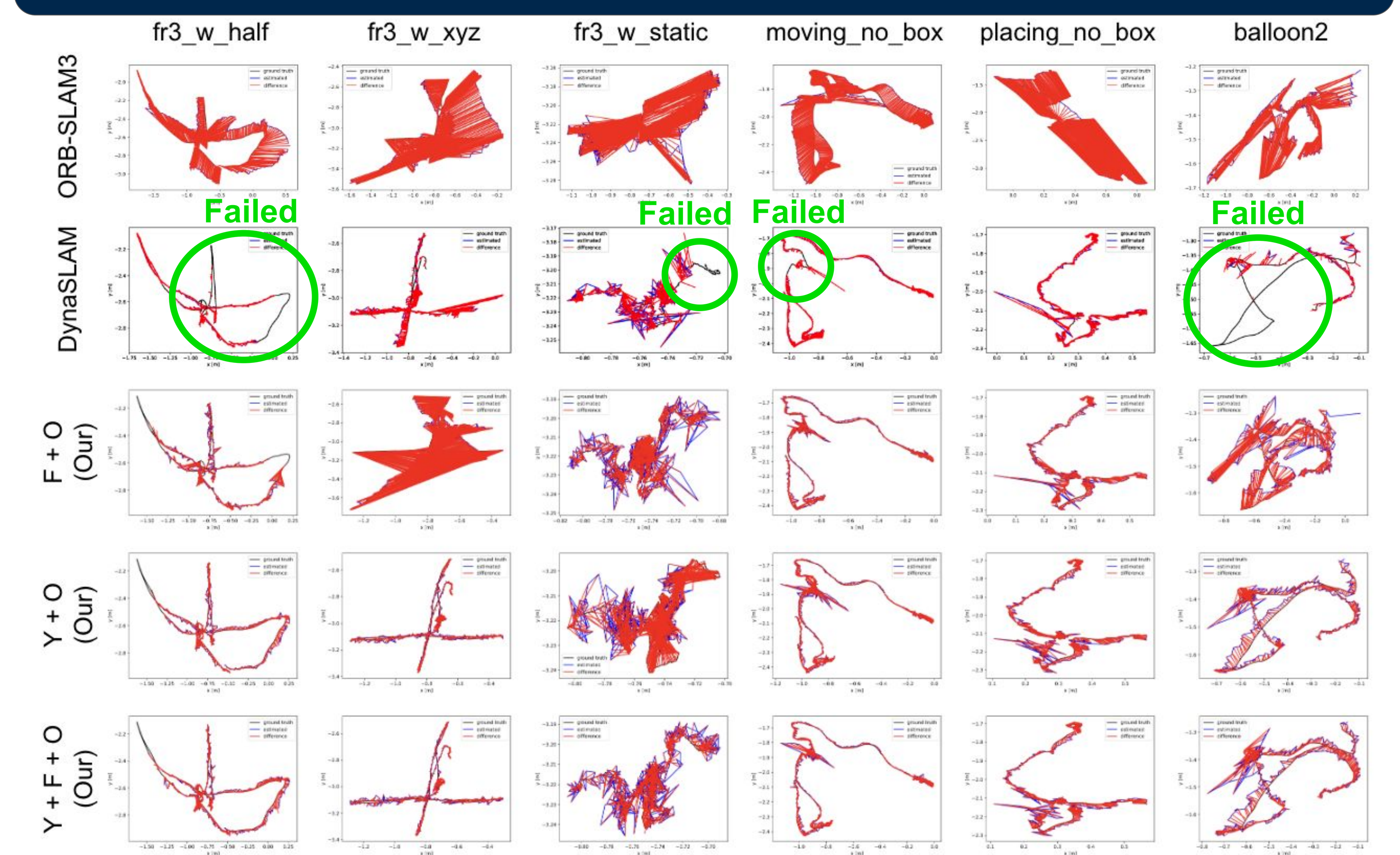


- **ORB Extraction**: Obtain the correspondences of ORB feature points between the current frame and the previous frame.
- **Remove Dynamic Feature Points**: Use the dynamic region mask to filter out the feature points on moving objects.
- **Localization**: Estimate the camera pose by tracking the remaining static feature points.

Results

	ORB-SLAM3	DynaSLAM	Y+O (Our)	F+O (Our)	Y+F+O (Our)
Platform (CPU + GPU)	i9 + RTX4060	i9 + Tesla M40	i9 + RTX4060	i9 + RTX4060	i9 + RTX4060
Inference Time [ms]	-	195	11.213	15.088	28.05
Tracking Time [ms]	10-16	429.63	11-17	11-17	11-17
Real-time	✅	❌	✅	✅	✅

Results



Dataset	ORB-SLAM3	F+O (Our)		Y+O (Our)		Y+F+O (Our)	
	ATE [m]	ATE [m]	Improvement [%]	ATE [m]	Improvement [%]	ATE [m]	Improvement [%]
fr3_w_half	0.332	0.043	86.99%	0.031	90.66%	0.031	90.54%
fr3_w_xyz	0.427	0.378	11.41%	0.018	95.82%	0.017	95.94%
fr3_w_static	0.291	0.011	96.10%	0.015	94.82%	0.008	97.34%
moving_no_box	0.179	0.030	83.24%	0.049	72.76%	0.032	81.89%
placing_no_box	0.772	0.032	95.85%	0.031	96.04%	0.027	96.49%
balloon2	0.227	0.114	49.70%	0.032	85.70%	0.043	80.96%

- **ORB-SLAM3 [1]** shows inaccurate tracking in dynamic scenes.
- **DynaSLAM [4]** often failed tracking and cannot run in real-time.
- **FastSAM/YOLO11 + Optical Flow (Our)** perform well in simple motion but struggles with complex dynamics.
- **YOLO11 + FastSAM + Optical Flow (Our)** achieves the lowest or second-lowest ATE performance in most sequences, with robust, reliable tracking under occlusion and motion.

Our Approach Achieves High Performance with Real-Time Capability!!!

Future Works

- **Apply in Real-World Scenario**: Test our method in the real-world environments and validate its robustness and performance in real-time.
- **Mapping and Semantic**: Render the map and integrate semantic segmentation for better scene understanding.

References: [1] C. Campos, R. Elvira, J. J. G. Rodr´ıguez, J. M. Montiel, and J. D. Tard´os, "ORB-SLAM33: An accurate open-source library for visual, visual–inertial, and multimap slam," IEEE transactions on robotics, vol. 37, no. 6, pp. 1874–1890, 2021.
[2] X. Zhao, W. Ding, Y. An, Y. Du, T. Yu, M. Li, M. Tang, and J. Wang, "Fast segment anything," 2023.
[3] G. Jocher and J. Qiu, "Ultralytics YOLO11," 2024.
[4] B. Bescos, J. M. F´acil, J. Civera, and J. Neira, "Dynaslam: Tracking, mapping, and inpainting in dynamic scenes," IEEE robotics and automation letters, vol. 3, no. 4, pp. 4076–4083, 2018

Acknowledgment: We would like to thank Prof. Maani Ghaffari, Dr. Minghan Zhu, and the GSIs for their inspiring lectures, helpful instruction, and support during the semester and throughout this project.