DetectFusion: Detecting and Segmenting Both Known and Unknown Dynamic Objects in Real-time SLAM

Ryo Hachiuma†, Christian Pirchheim‡, Dieter Schmalstieg‡, Hideo Saito †
† Keio University, ‡Graz University of Technology
†{ryo-hachiuma, hs}@keio.jp, ‡{pirchheim, schmalstieg}@icg.tugraz.at

Abstract

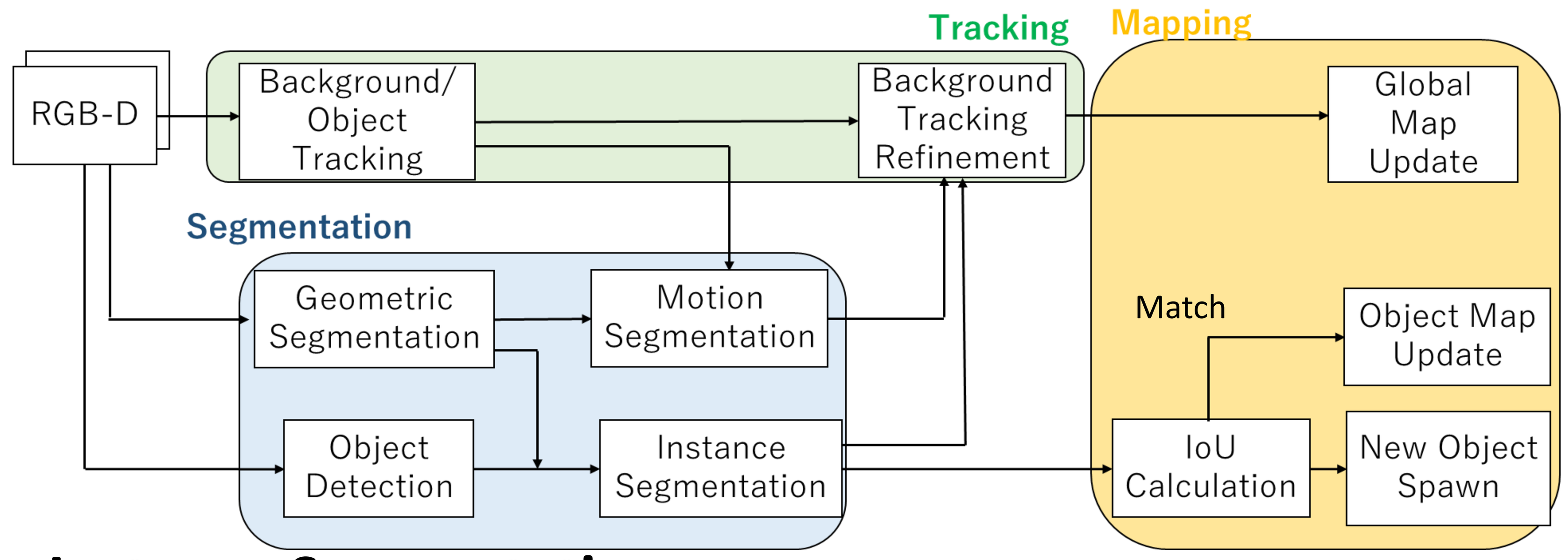
Simutaneous Localization and Mapping (SLAM) jointly estimates a 3D map of an unknown environment and the pose of an RGB/RGB-D camera. Most contemporary SLAM systems (1) assume that the scene is static and (2) do not assign object semantics to their 3D maps!

Approaches for SLAM in dynamic environments

- reconstruct non-rigid objects (e.g., a single moving person)
- reconstruct only the static background and ignore dynamic objects
- reconstruct multiple rigid dynamic objects (our approach)

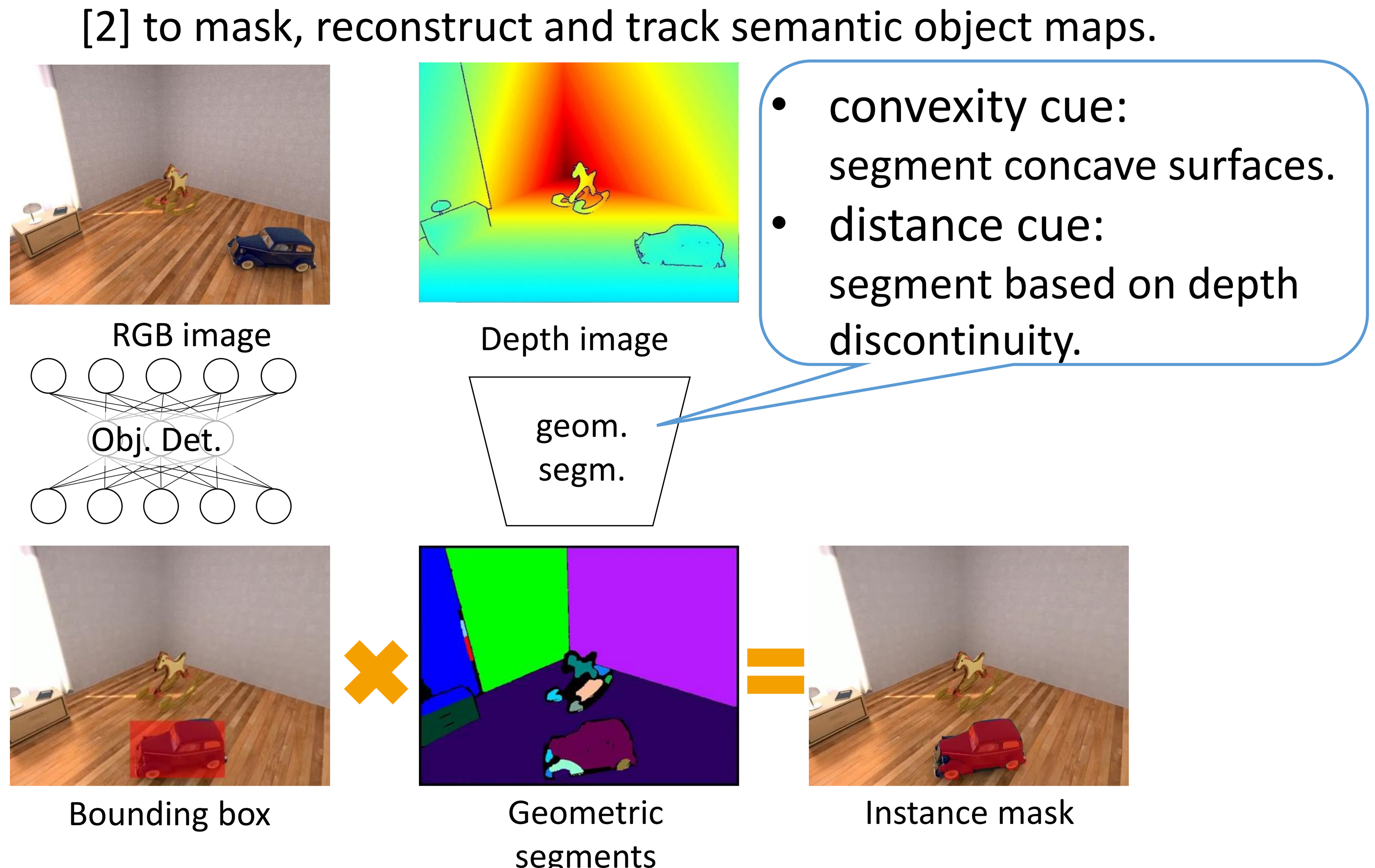
The key system component of our approach is the generation of instance object masks to reconstruct and track multiple semantic object maps independently, in addition to the static background map.
Our RGB-D SLAM **runs in real time** and can robustly **handle semantically known and unknown objects** that can move dynamically in the scene.

Method



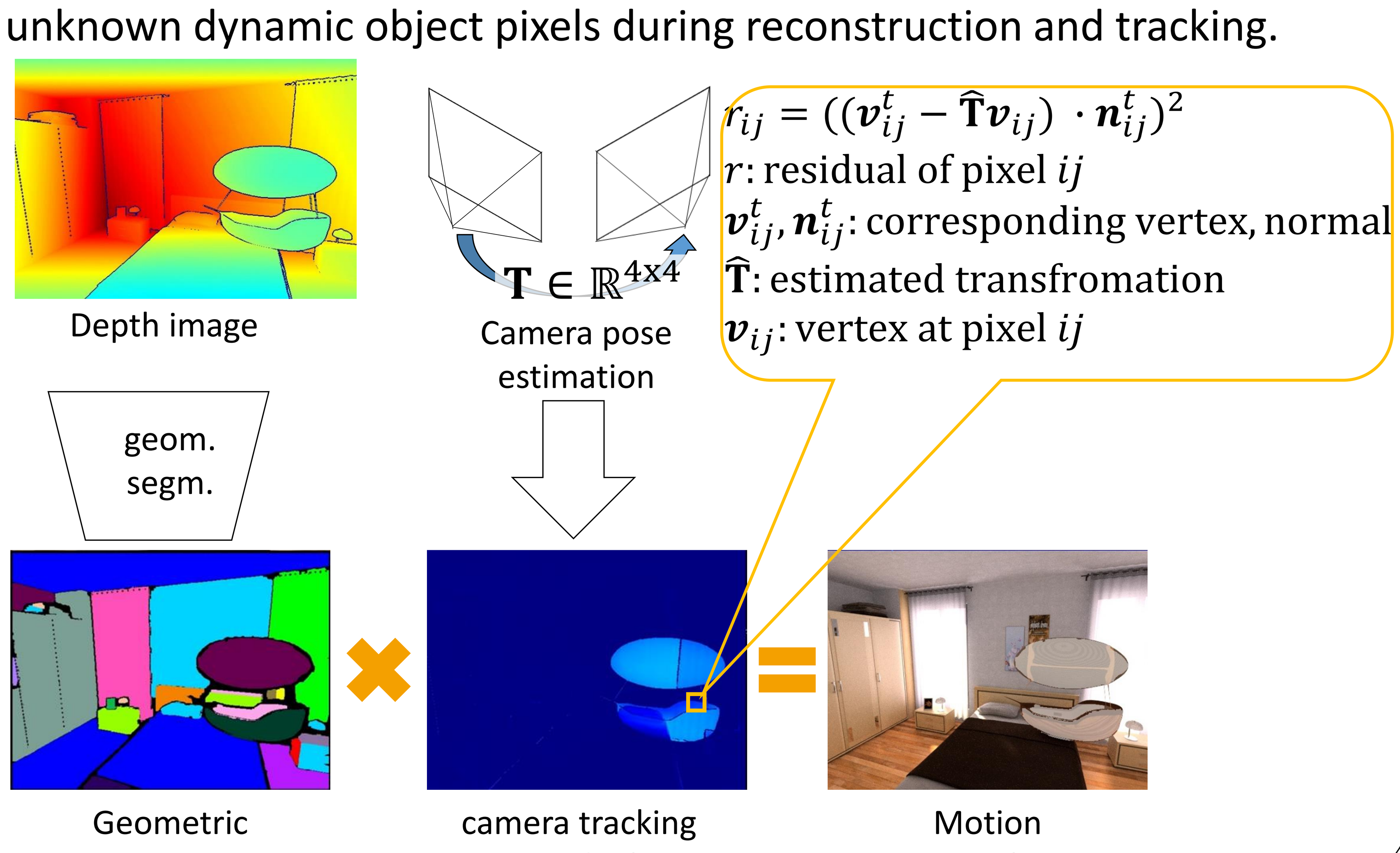
1. Instance Segmentation

Combine object detection [1] and geometric segmentation [2] to mask, reconstruct and track semantic object maps.



2. Motion Segmentation

Combine the geometric segmentation and ICP residual heatmap to ignore unknown dynamic object pixels during reconstruction and tracking.



Contributions

Instance segmentation of known objects at full-frame rate

- We achieve real-time performance on semantic instance segmentation by a parallel computation approach on CPU and GPU (detect-while-segment).
- Mask R-CNN performs sequential computation (detect-then-segment).

Motion segmentation of unknown moving objects

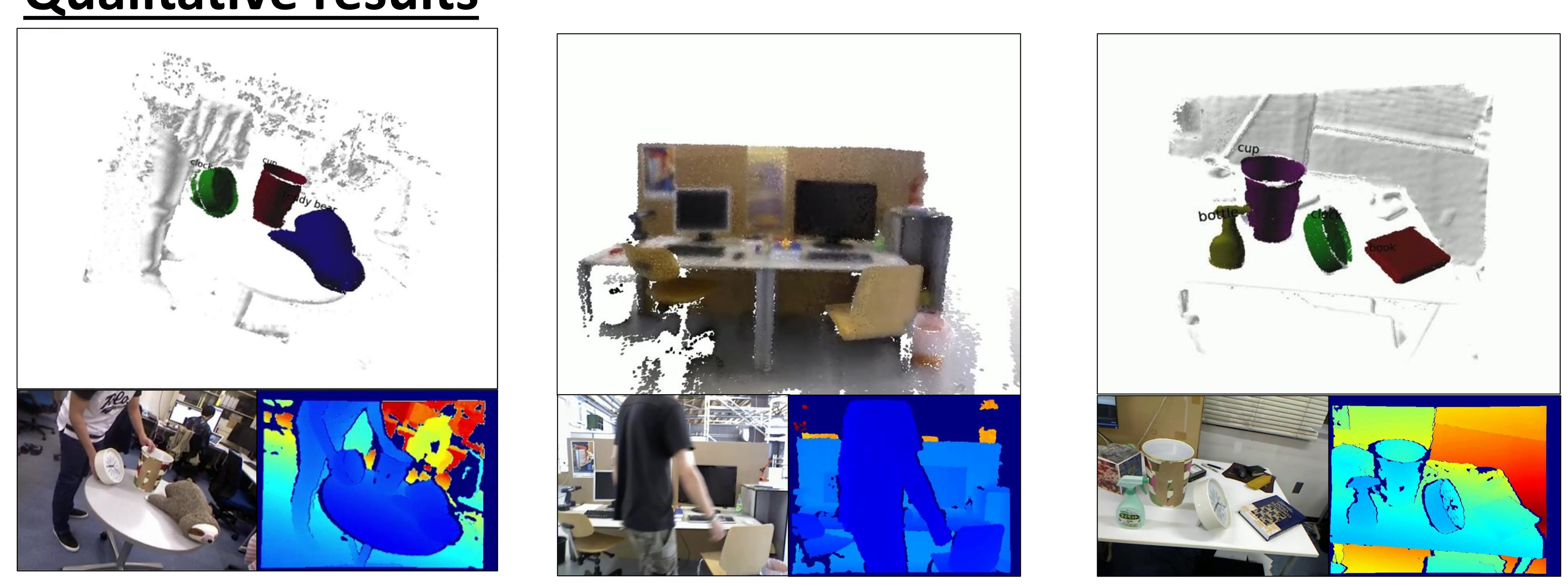
- We propose a method for detecting and segmenting the motion of semantically unknown objects and ignore them explicitly, thus further improving the accuracy of camera tracking and reconstruction.
- In MaskFusion [6], unknown moving objects may disturb camera tracking.

Comparison

System	Static Map	Object Maps	Instance Segmentation	Motion Segmentation	Real-time
ElasticFusion [3]	✓				✓
StaticFusion [4]	✓			✓	✓
CoFusion [5]	✓	✓		✓	✓
MaskFusion [6]	✓	✓	keyframe-rate		▲
MID-Fusion [7]	✓	✓	Offline	✓	
Ours	✓	✓	frame-rate	✓	✓

Result

Qualitative results



Our RGB-D SLAM reconstruct multiple moving objects with semantic labels. Non-rigid objects are ignored from tracking and mapping.

Quantitative results

1. Tracking performance

We evaluated the tracking accuracy on TUM dataset. Our method outperformed dynamic SLAM methods on the **highly dynamic sequences**.

Setting	Sequence	ATE RMSE (cm) ↓				
		EF[3]	SF[4]	CF[5]	MF[6]	Ours
Slightly Dynamic	f3s_static	**0.9**	1.3	1.1	2.1	1.5
	f3s_xyz	**2.6**	4.0	2.7	3.1	5.2
	f3s_sphere	13.8	4.0	**3.6**	5.2	4.1
Highly Dynamic	f3w_static	6.2	1.4	55.1	3.5	3.6
	f3w_xyz	21.6	12.7	69.6	10.4	**8.5**
	f3w_sphere	20.9	39.1	80.3	10.6	**7.2**

2. Runtime performance

Our method achieved **22 FPS** while performing segmentation in every frame. (GTX1080Ti, Core i7-6950X 3.0GHz CPU, 64GB RAM)

Component	Runtime [ms]
Initial Tracking	3.80 per model
Geometric segmentation	6.36
Object Detection	19.1
Motion Segmentation	2.72
Object Mask Generation	0.57
Camera pose Refinement	6.30
Mapping	2.05 per model
Total	46.1 + 5.16 per model

The total runtime depends on the number of objects in the scene. (e.g. 22FPS for 0 obj., 19 FPS for 1 obj.)

References

[1] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement.CoRR , abs/1804.02767, 2018.
[2] K. Tateno, F. Tombari, and N. Navab. Real-time and scalable incremental segmentation on dense SLAM. In IROS, 2015.
[3] Thomas Whelan, Renato F Salas-Moreno, Ben Glocker, Andrew J Davison, and Stefan Leutenegger. Elasticfusion: Real-time dense SLAM and light source estimation. In IJRR, 2016.
[4] R. Scona, M. Jaimez, Y. R. Petillot, M. Fallon, and D. Cremers. Staticfusion: Background reconstruction for dense RGB-D SLAM in dynamic environments. In ICRA, 2018.
[5] Martin Rünz and Lourdes Agapito. Co-fusion: Real-time segmentation, tracking and fusion of multiple objects. In ICRA, 2018.
[6] Martin Rünz, Maud Buffier, and Lourdes Agapito. Maskfusion: Real-time recognition, tracking and reconstruction of multiple moving objects. In ISMAR 2018.
[7] Binbin Xu, Wenbin Li, Dimos Tzoumanikas, Michael Bloesch, Andrew J. Davison, and Stefan Leutenegger. Mid-fusion: Octree-based object-level multi instance dynamic SLAM, In ICRA, 2019.