筑波大学
University of Tsukuba
大学院 知能機能システム専攻
Department of Intelligent Interaction Technologies

AIST

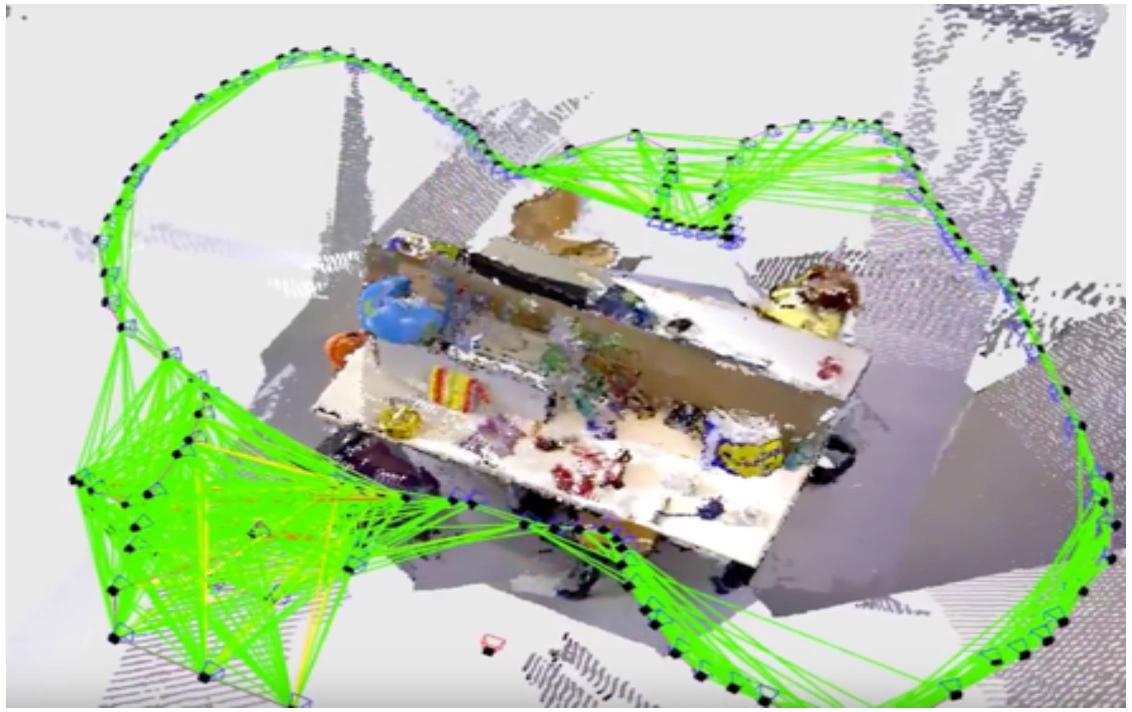
Multi-Purpose SLAM Framework for Dynamic Environment

Sun, Leyuan (Univ. of Tsukuba, National Inst. of AIST)
Kanehiro, Fumio (National Inst. of AIST)
Kumagai, Iori (National Inst. of AIST)
Yoshiyasu, Yusuke (CNRS-AIST JRL)

2020 IEEE/SICE International Symposium on System Integration
Hawaii Convention Center, Honolulu, Hawaii, USA
January 12 - 15, 2020

What is SLAM

- Simultaneous Localization and Mapping (SLAM)



Dense visual SLAM[1]

- Assumption: Static environment

- Estimate camera pose by keypoints sets in continuous frames

Keypoints matching results

[1] Kerl, Christian, Jürgen Sturm, and Daniel Cremers. "Dense visual SLAM for RGB-D cameras." *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013.

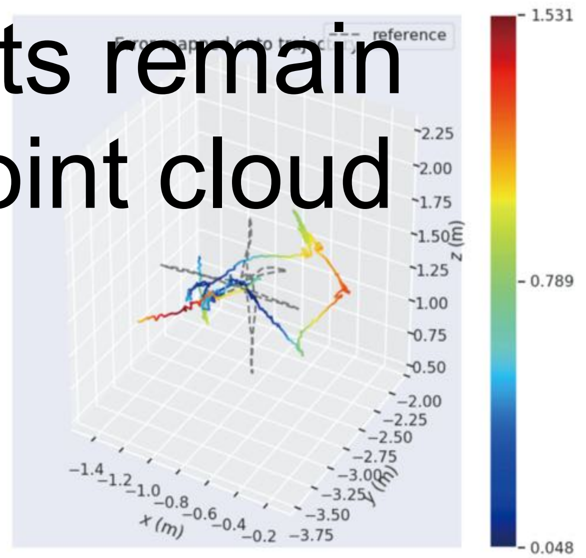
Issues need to be solved when SLAM in dynamic environment

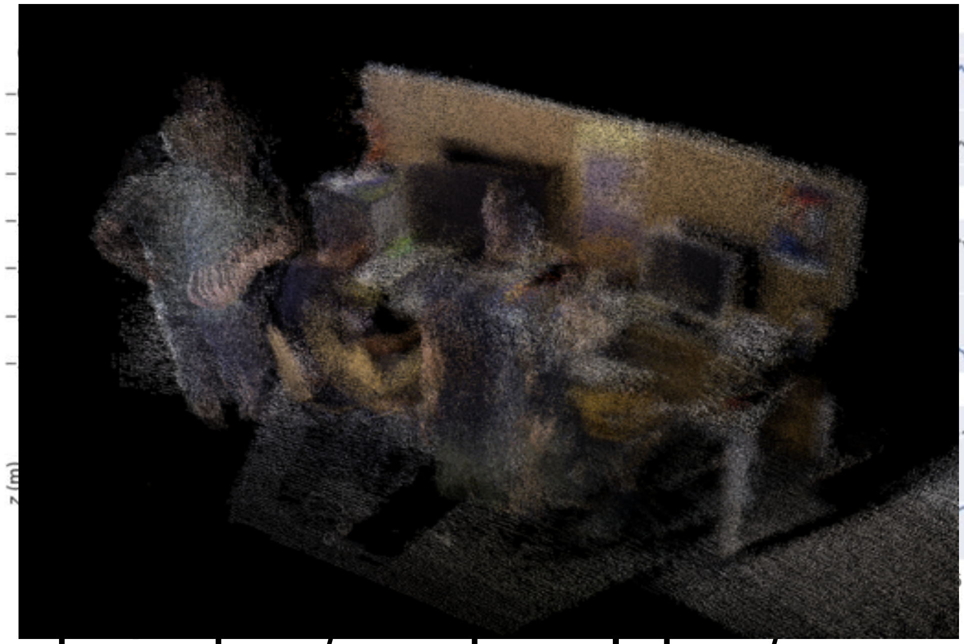
- Inaccurate tracking

TUM RGB-D dynamic sequence fr3_walking_xyz

Keypoints matching in dynamic environment

- Dynamic objects remain in the dense point cloud map



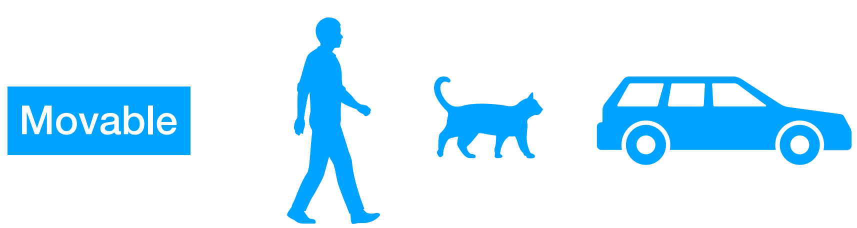


Camera trajectory/ xyz translation/ yaw.pitch.roll

Dense point cloud with dynamic objects

Multi-purpose for SLAM in Dynamic Environment

- Dynamic including **movable** and **moving**



- Movable could be static
- Moving objects could be these moved by moveable objects such as: 

- To obtain **accurate tracking**
    - Moving objects need to be segmented out
- To obtain **dense point cloud map** of static objects
    - Movable objects need to be segmented out
- To do **online processing**
    - Segmentation need be processed on real-time

Related work

Categories	Real-time	Non-realtime
Movable	Dynamic-SLAM[1]	-
Moving	Reference[2]	Reference[3]
Both	-	DynaSLAM[4]

- **Real-time(I) for purpose of online processing;**
- **Moving(II) segmentation for purpose of accurate tracking;**
- **Movable + moving(III) segmentation for purpose of dense point cloud;**

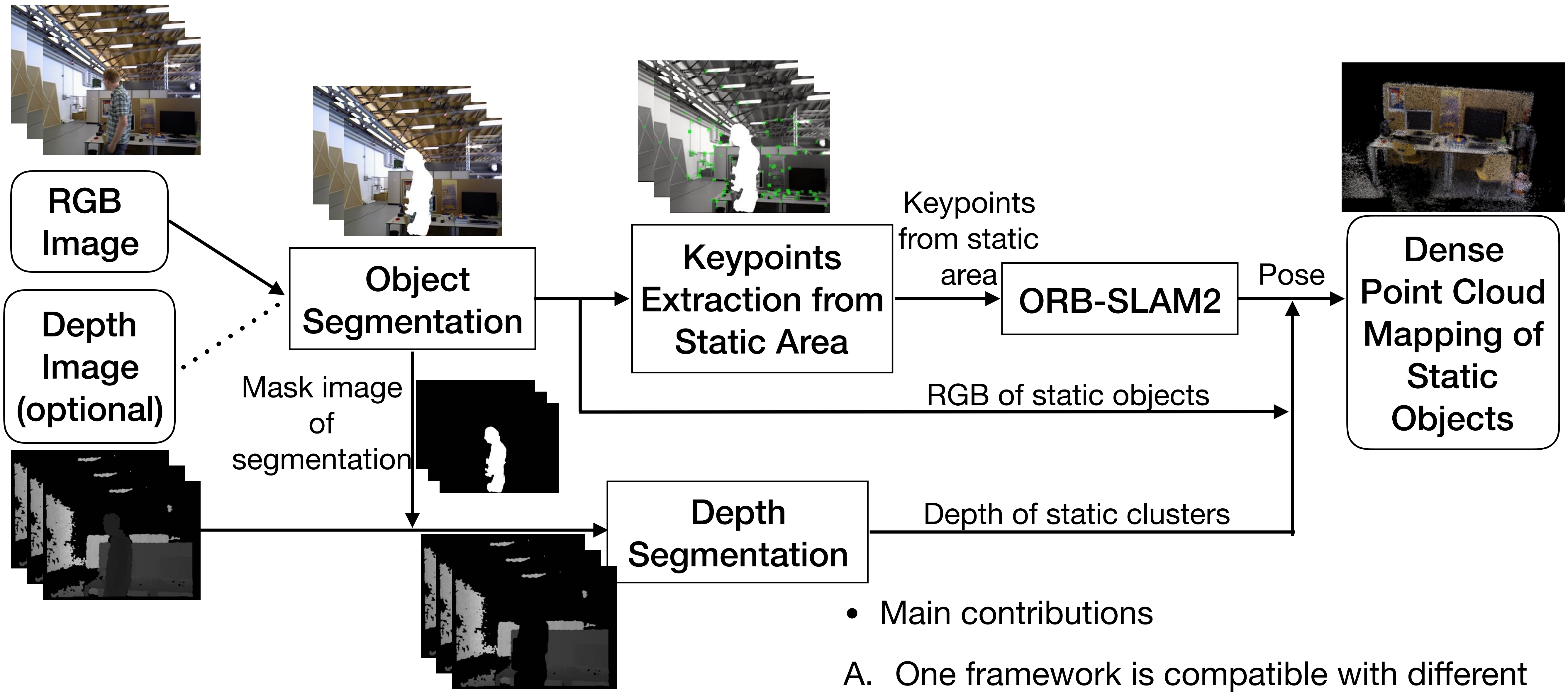
I ∩ II ∩ III = real-time+moving+movable;

- **Not exist**
- **Not suitable for accurate tracking**

- Conclusion: **Difficulty for using one method for all purposes**
- **Purpose of this research:**
  - Propose a **multi-purpose** SLAM framework which is configurable depending on user's purpose
- **Major features**
  - Useful to **compare** different segmentation methods on a single platform
  - Generate **dense point cloud map** of static objects if depth information is available

[1] Xiao, Linhui, et al, "Dynamic-SLAM: Semantic monocular visual localization and mapping based on deep learning in dynamic environment," *Robotics and Autonomous Systems*, vol. 117, 2019, pp. 1-16.
[2] Cheng, Jiyu, Yuxiang Sun, and Max Q-H. Meng, "Improving monocular visual SLAM in dynamic environments: an optical-flow-based approach," *Advanced Robotics* vol. 33, no.12, pp. 576-589
[3] Wang, Runzhi, et al, "A New RGB-D SLAM Method with Moving Object Detection for Dynamic Indoor Scenes," *Remote Sensing*, vol. 11, no.10, 2019, pp. 1143.
[4] Bescos, Berta, et al. "DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes," *IEEE Robotics and Automation Letters* vol. 3, no.4, 2018, pp. 4076-4083.

Overview of framework



- Main contributions

A. One framework is compatible with different segmentation methods for different purposes

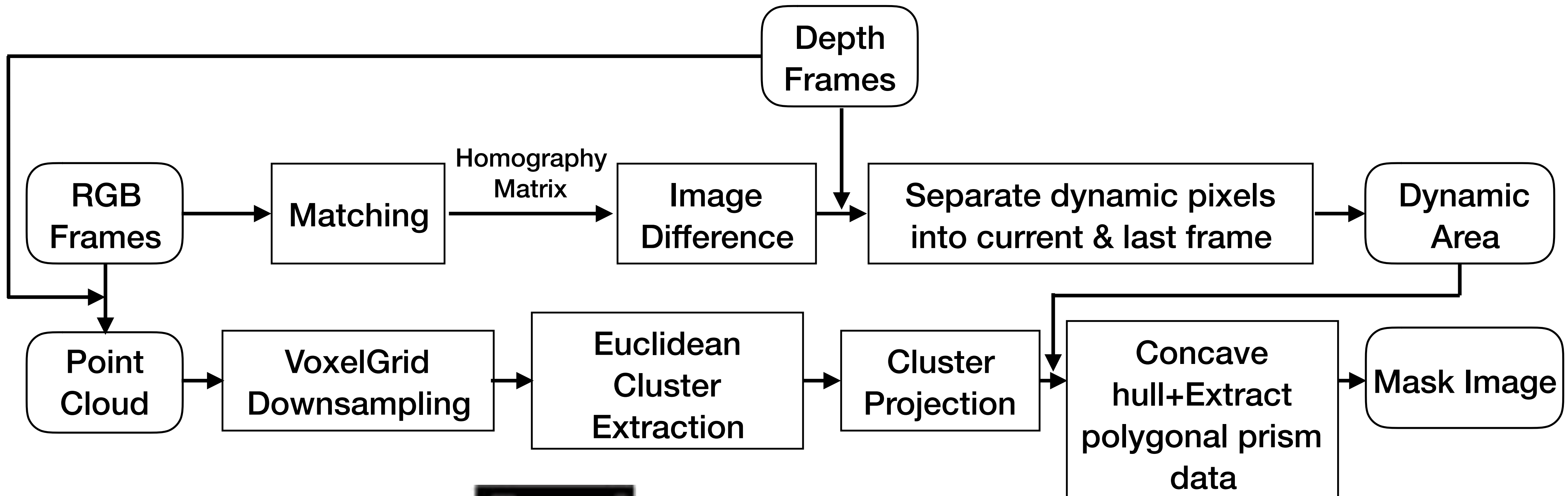
B. Uniformly distributed keypoints extraction in static area

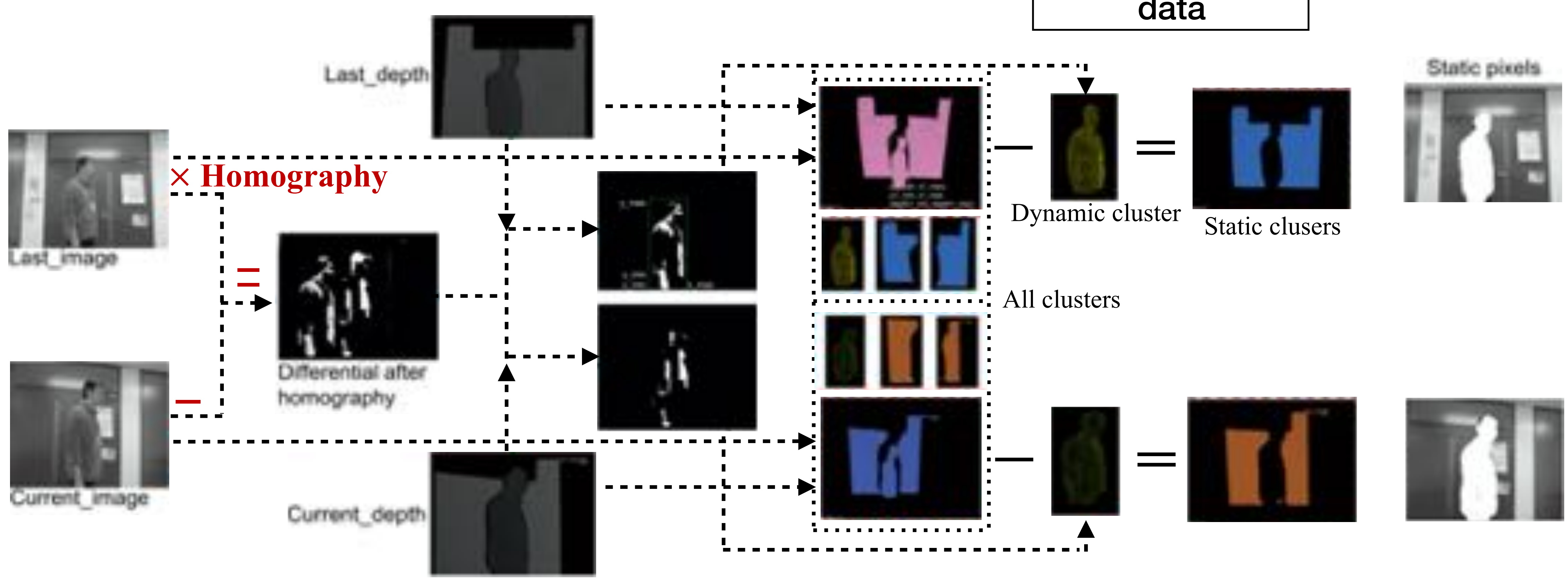
C. Dense point cloud map of static objects

[1] Mur-Artal, Raul, and Juan D. Tardós, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Transactions on Robotics*, vol. 33, no.5, 2017, pp. 1255-1262.

Segmentation

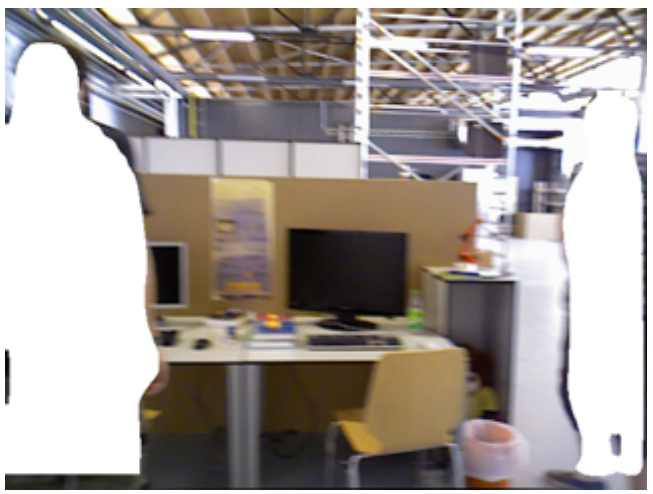
- Geometry-based method for moving object



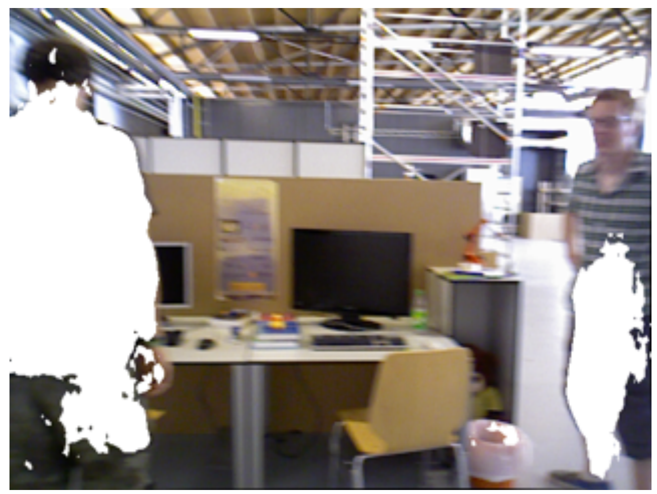


Segmentation

- Deep learning-based method for **movable** objects



Mask R-CNN[1]



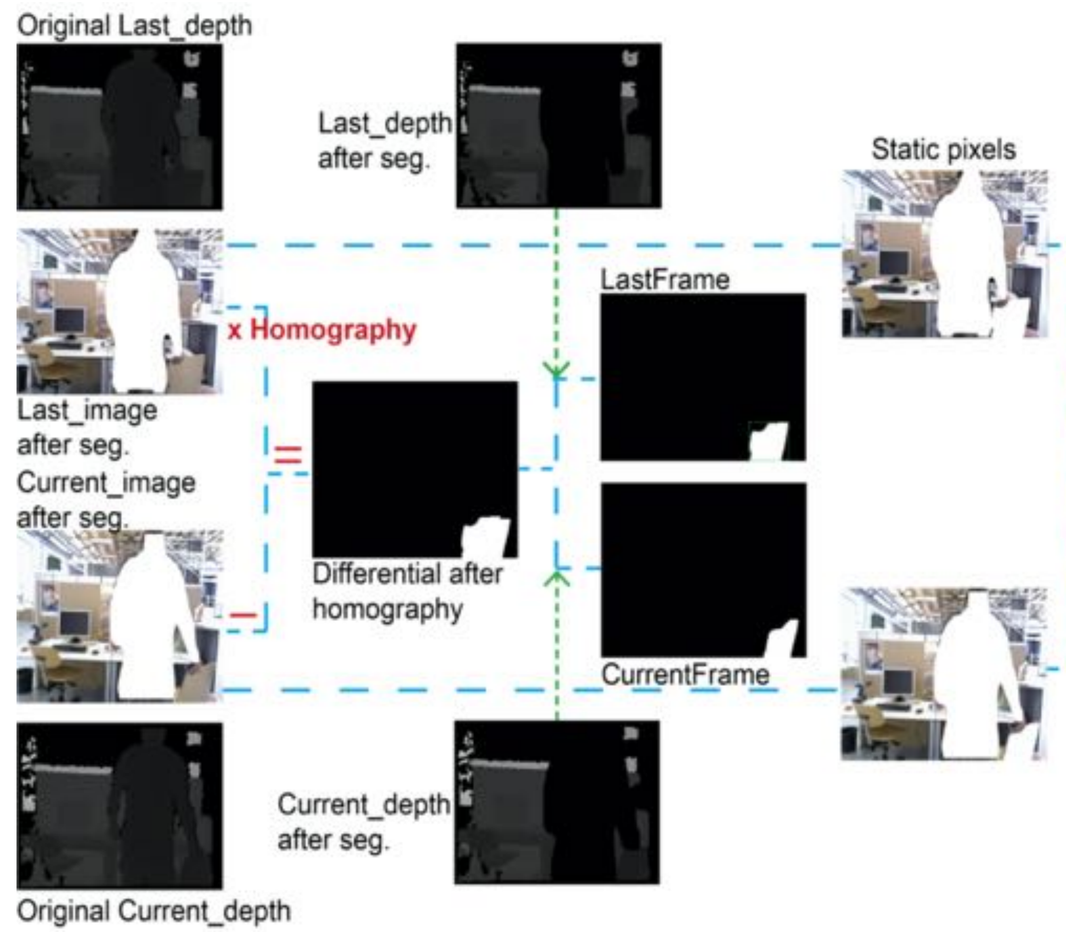
Lightweight model deep learning(LWDL)[2]

Table. Computational cost comparison

Method	Mask R-CNN	LWDL
Processing Time [ms/frame]	4137(CPU) 673 (GPU)	167 (CPU) 17 (GPU)

CPU: i7-7700hq 2.8GHz/ GPU: Nvidia GTX 1060 6GB

- Combination for **movable and moving** objects



Mask R-CNN + Geometry-based

[1] He, Kaiming, et al, "Mask R-CNN," *The IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2961-2969
[2] https://github.com/AntiAegis/Human-Segmentation-PyTorch

Dilated Mask

Many outliers on the contour of mask

Segmentation mask, keypoints matching result

Filter outliers

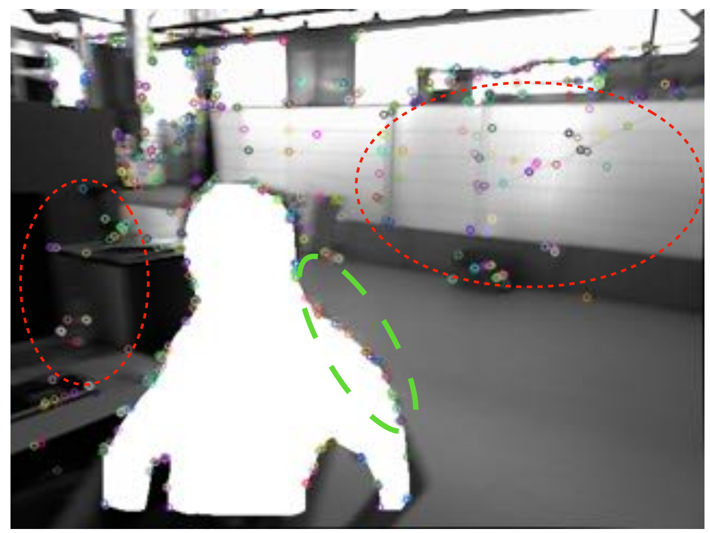
Dilated mask, keypoints matching result

When mask cannot cover objects accurately

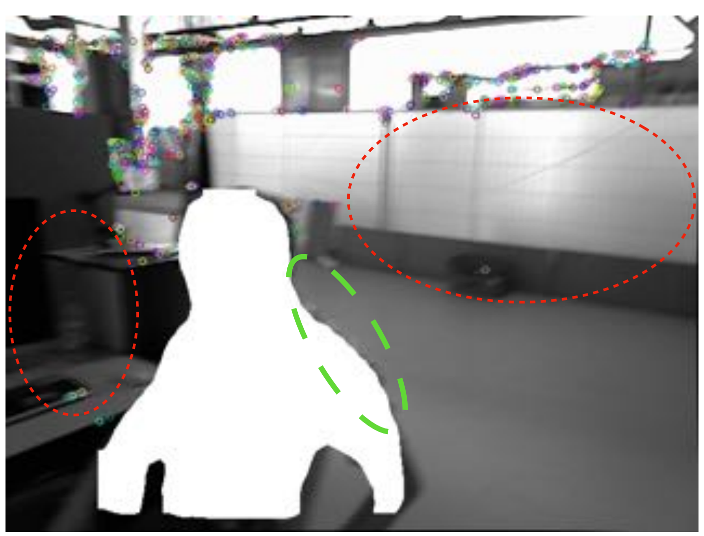
Make up the inaccuracy of segmentation's contour

Extraction of uniformly distributed keypoints

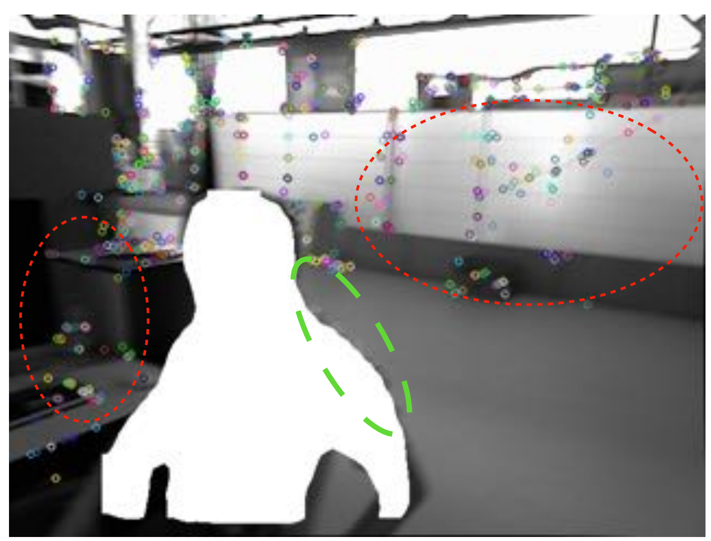
From fr3_sitting_halfsphere
Bad illumination and **blurred**



From fr3_sitting_halfsphere
ORB-SLAM2 extractor



From fr3_sitting_halfsphere
OpenCV extractor



Improved OpenCV extractor
Through 30*30 patch

Keypoints on contours or not

Keypoints uniformly distributed or not

Table. Absolute Trajectory Error(ATE) [1] comparison between 3 extractors by TUM dataset

Sequence unit/m	ORB-SLAM2	OpenCV
fr1_xyz	0.0135	0.0120
fr1_desk	**0.0193**	0.0219
fr2_3h	0.1128	**0.0928**
fr2_desk	0.0072	0.0145
fr3_loh	**0.0098**	0.0159
fr3_sh	**0.0310**	0.1958

Accuracy : Improved OpenCV ≈ ORB-SLAM2 > OpenCV

Table. Processing time comparison

Processing time [s/frame]	Keypoint extraction	Tracking
ORB-SLAM2 extractor	0.00124	0.038
Improved OpenCV extractor	0.00132	0.043

Real-time ability

Depth segmentation and Point cloud of static objects


Original depth

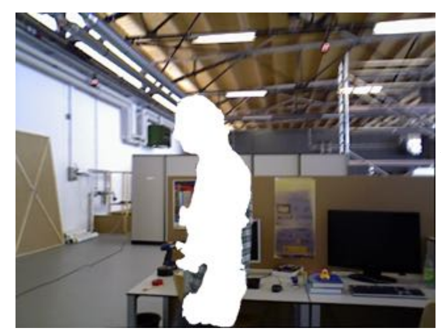

Binary mask


Segmented depth

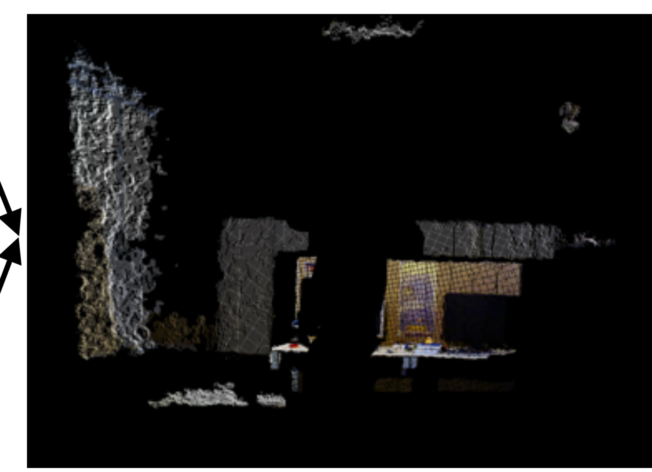

Original RGB


Original depth


Dense point cloud
with dynamic objects

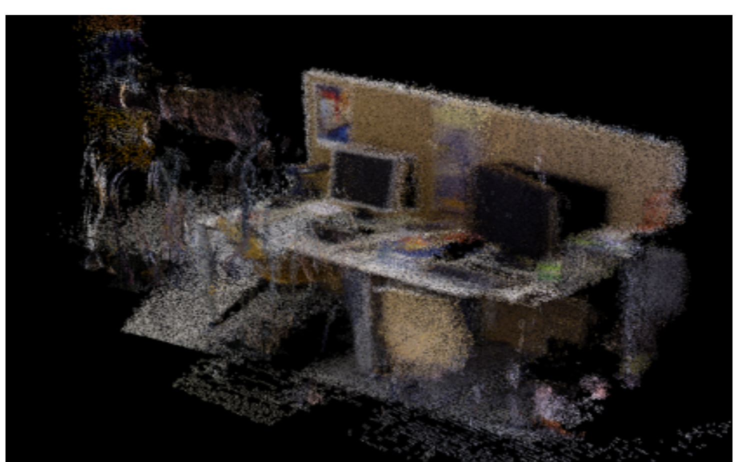

Segmented RGB


Segmented depth


Dense point cloud
of static objects

Experiment

TUM RGB-D Datasets Sequence:	Real-time Movable Segmentation	ATE (cm) RMSE			Translation RMSE (cm/frame)			Rotation RMSE (deg/frame)		
		Dynamic-SLAM [1]	ORB-SLAM2 with LWDL	Improvement	Dynamic-SLAM [1]	ORB-SLAM2 with LWD	Improvement	Dynamic-SLAM [1]	ORB-SlAM2 with LWDL	Improvement
Low Dynamic Environment	**fr2/desk_with_person**	1.873	0.316(Traj% 85.62)	**83.12%**	1.958	0.422	**78.45%**	0.833	0.283	**66.03%**
	fr3/sitting_xyz	0.601	0.298(Traj% 95.56)	**50.42%**	0.998	0.517	**48.20%**	0.613	0.325	**46.98%**
	fr3/sitting_halfsphere	1.461	0.992(Traj% 71.08)	**32.10%**	1.451	0.828	**42.94%**	0.551	0.439	**20.32%**
	fr3/sitting_rpy	3.448	4.023(Traj% 61.08)	-16.68%	4.303	4.832	-12.29%	0.991	1.231	-24.22%
High Dynamic Environment	**fr3/walking_xyz**	1.324	2.021(Traj% 98.49)	-52.64%	1.796	0.970	**45.99%**	0.598	0.526	**12.04%**
	fr3/walking_halfsphere	2.139	0.434(Traj% 86.79)	**79.71%**	2.192	0.813	**62.91%**	0.666	0.408	**38.74%**
	fr3/walking_rpy	6.025	0.669(Traj% 81.10)	**88.90%**	5.605	0.544	**90.29%**	1.149	0.429	**62.66%**



Dense point cloud map

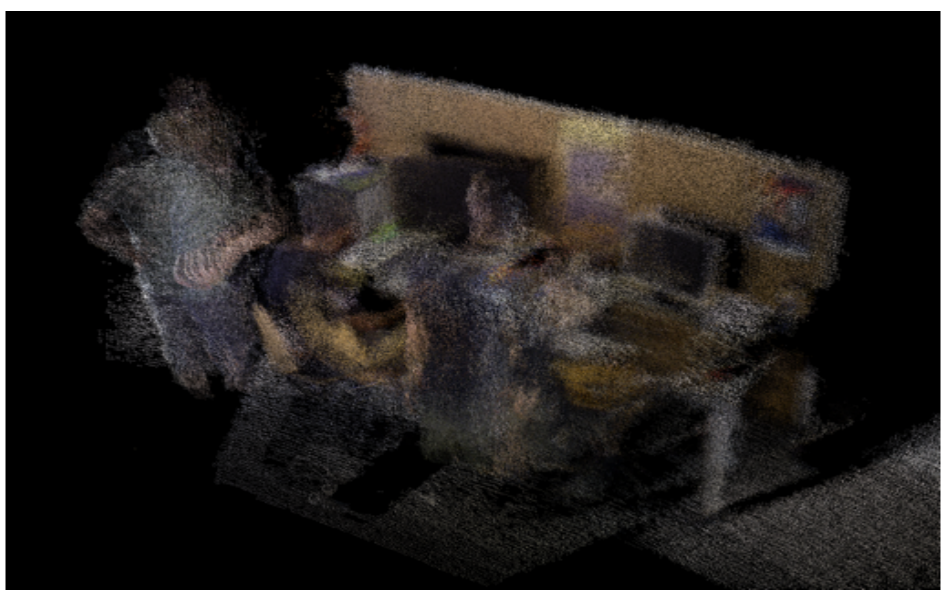
- **More accurate than Dynamic-SLAM [1]**
- **Our framework could obtain dense point cloud map**

[1] Xiao, Linhui, et al, "Dynamic-SLAM: Semantic monocular visual localization and mapping based on deep learning in dynamic environment," *Robotics and Autonomous Systems*, vol. 117, 2019, pp. 1-16.

Experiment

- **Non-realtime for moving objects:**

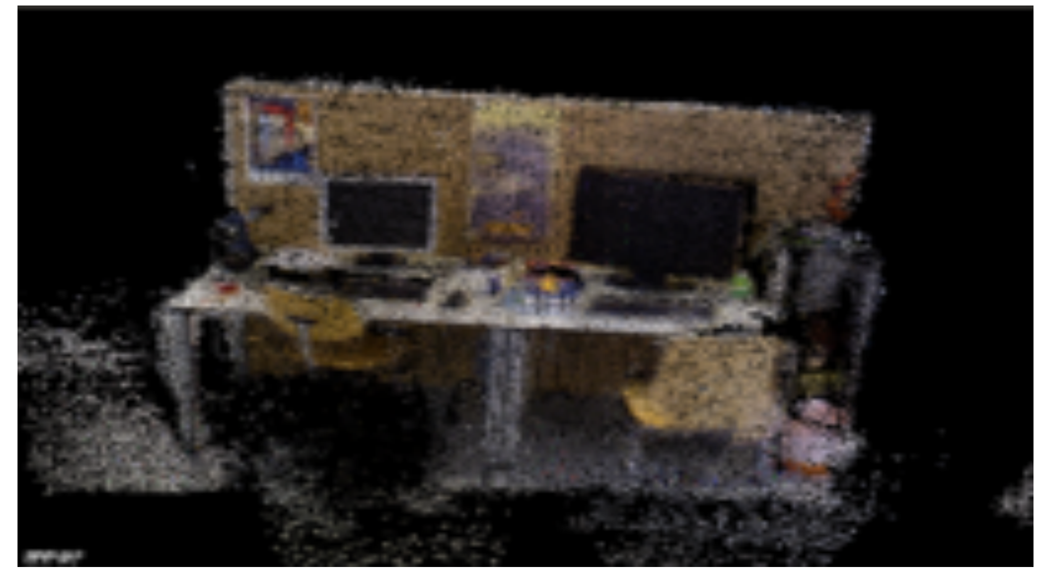
Sequence	ATE (m) RMSE		RPE Translation(m/s) /Rotation(deg/s) RMSE	
	Reference[1]	GS	Reference[1]	GS
fr3_sittting_static	0.0066	0.0052	0.0077/0.2595	0.0058/0.2342
fr3_sitting_halfsphere	0.0196	0.0185	0.0245/0.5643	0.0211/0.6023
fr3_walking_static	0.3080	0.5532	0.1881/3.2101	0.3011/4.5342



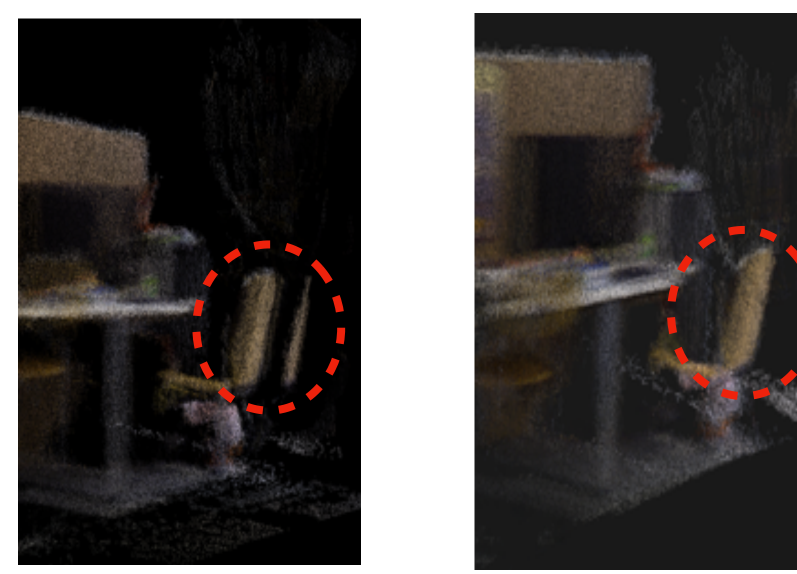
Dense point cloud map by GS

- **More accurate in some low dynamic environments**
- **Dense point cloud map**

- **Combination for movable + moving objects:**



Dense point cloud map



Only Mask R-CNN / Mask R-CNN + GS

- **Suitable for 3D model of static environment offline with moving objects in short duration**

[1] Wang, Runzhi, et al, "A New RGB-D SLAM Method with Moving Object Detection for Dynamic Indoor Scenes," *Remote Sensing*, vol. 11, no.10, 2019, pp. 1143.

Summary

- Single segmentation method could not meet all requirements for all purposes.
- The proposed framework could be compatible with different segmentation methods.
- The framework has similar or better performance compared with existing methods in some aspects.

筑波大学
University of Tsukuba
大学院 知能機能システム専攻
Department of Intelligent Interaction Technologies

AIST

Thanks for listening.