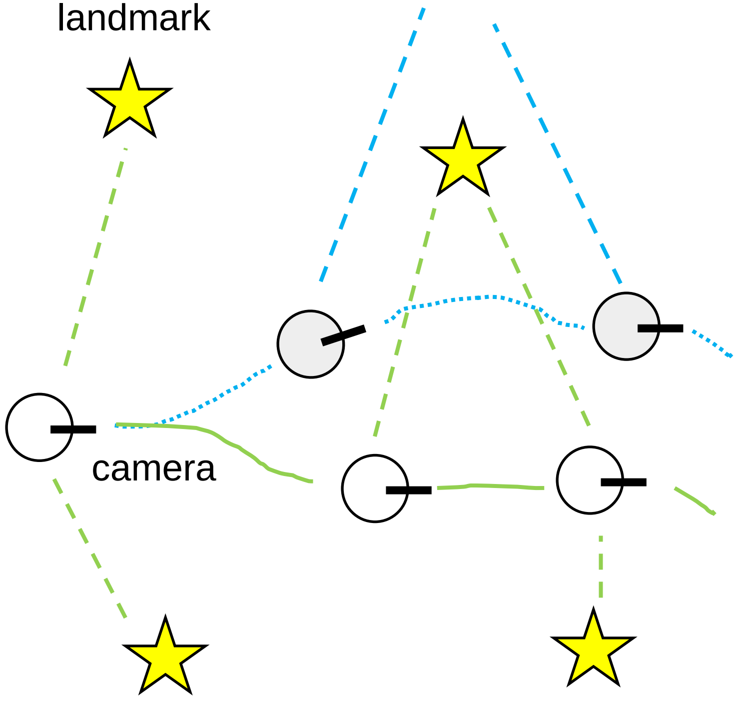
Structure SLAM with Points, Planes, and Objects

Benchun Zhou, Maximilian Gilles, and Yongqi Meng
Institute for Material Handling and Logistics (IFL),
Karlsruhe Institute of Technogy (KIT),
Karlsruhe, Germany

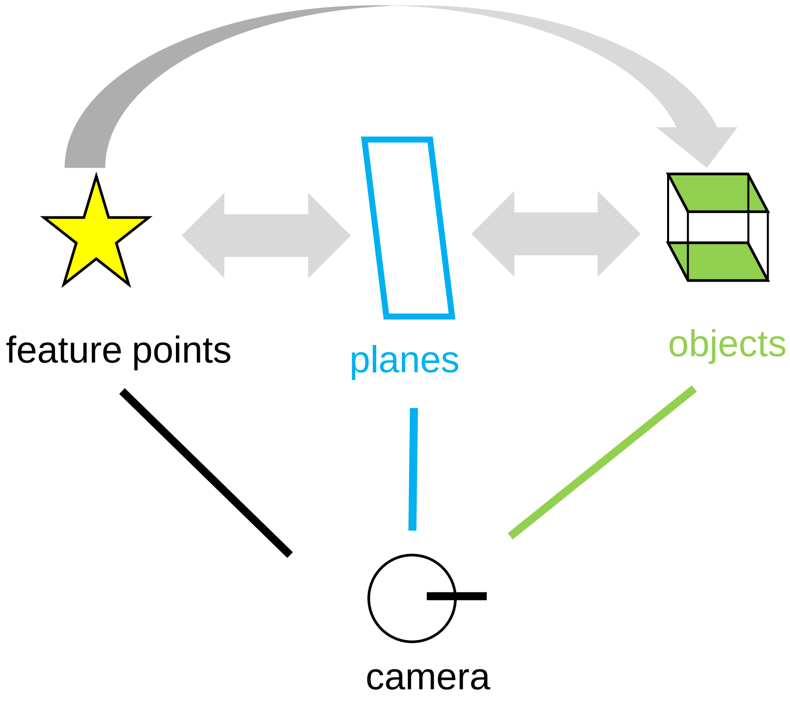
Contents

Background

- Research Question: Does the introduction of 3D objects can benefit robot localization?



Visual SLAM with feature points

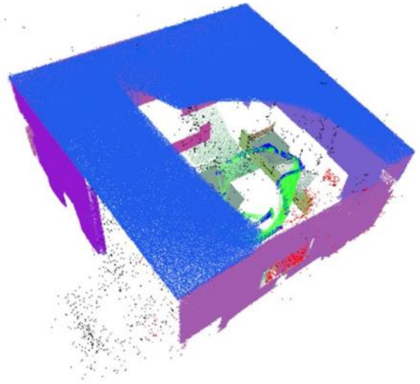


Visual SLAM with different landmarks

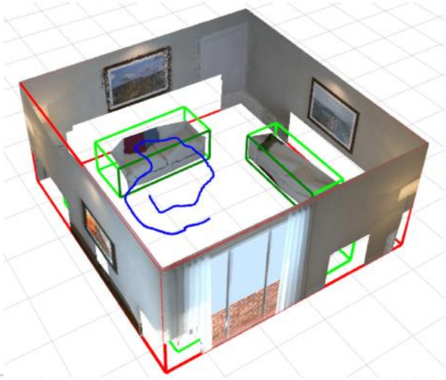
Literature Review

	Sensor	Point	Plane	Object	Data Association
\cite{davison2007monoslam}	RGB	image patch			patch search
\cite{klein2007parallel}	RGB	FAST			patch search
\cite{mur2015orb}	RGB	ORB			descriptor matching
\cite{hsiao2017keyframe}	RGB-D		plane		normal, distance, residual
\cite{zhang2019point}	RGB-D		plane		normal, distance, boundary
\cite{yang2016pop}	RGB	ORB	plane		normal, distance, polygon
\cite{li2021rgb}	RGB-D	ORB	plane		normal, distance
\cite{hosseinzadeh2017sparse}	RGB-D	ORB	plane		normal, distance
\cite{salas2013slam++}	RGB-D			3D model	feature matching
\cite{nicholson2018quadricslam}	RGB			quadric	not mentioned
\cite{liao2020rgb}	RGB-D			quadric	not mentioned
\cite{liao2022so}	RGB-D			quadric	not mentioned
\cite{yang2019cubeslam}	RGB	ORB		cuboid	in-object points
\cite{li2020view}	RGB			cuboid	geometric features
\cite{lin2021topology}	RGB			cuboid	geometric features, graph matching

Point-based SLAM
Mur-Artal et al, 2015



Point-plane SLAM
Zhang et al, 2019

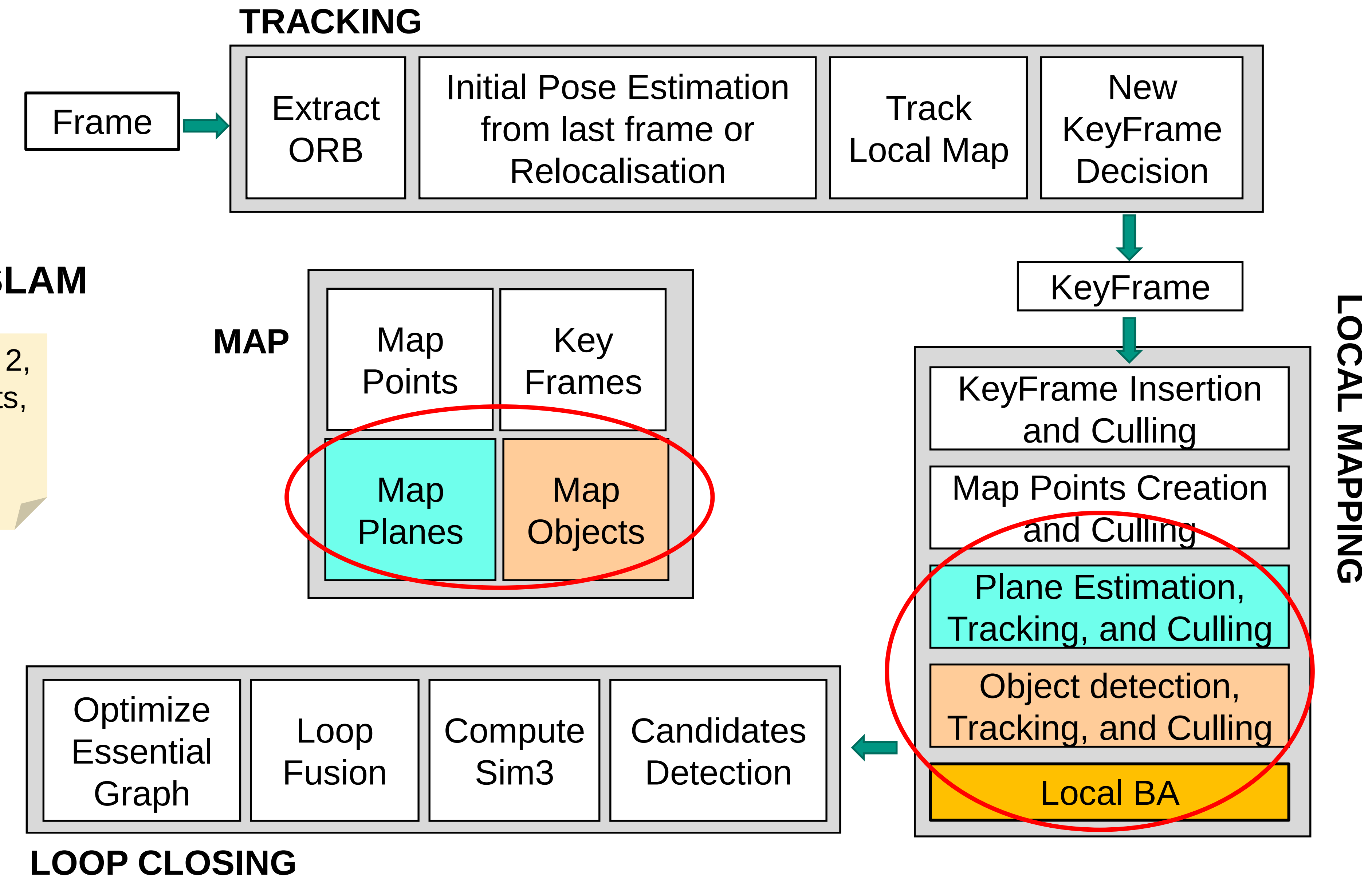


Point-plane-object SLAM
Yang et al, 2019

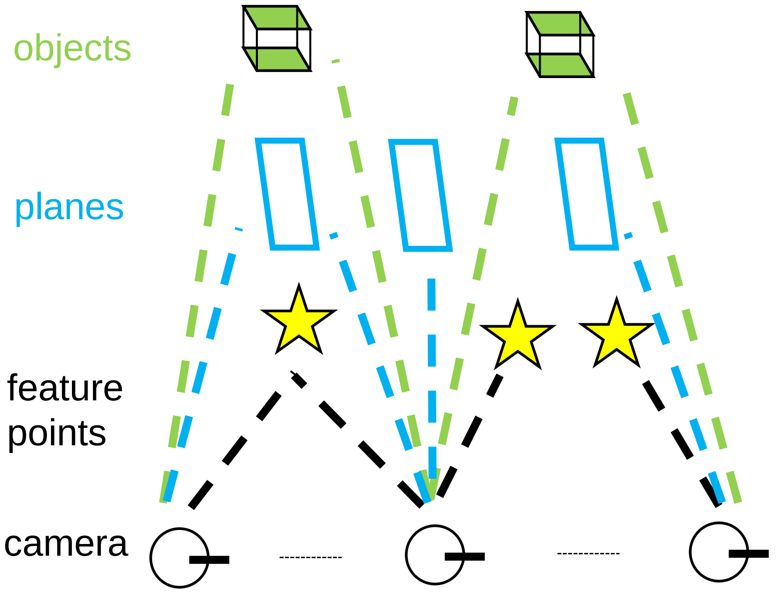
Framework

Point-plane-object SLAM

On the top of ORB SLAM 2, but add planes and objects, in detection, optimization and mapping process



Method: Landmark Detection and Association



- Representation
  - Camera pose $\boldsymbol{T_c} \in SE(3)$
  - Features point $\boldsymbol{P} = (x, y, z, 1)^T \in R^3$
  - Plane $\boldsymbol{\pi} = (n_x, n_y, n_z, d)^T \in R^3$
  - Object $\boldsymbol{O} = (\boldsymbol{T_o}, \boldsymbol{D})^T, \boldsymbol{T_o} \in SE(3), \boldsymbol{D} \in R^3$

$$SE(3) = \left\{ \boldsymbol{T} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix} \in R^{4\times4} \middle| \boldsymbol{R} \in R^{3\times3}, \boldsymbol{t} \in R^3 \right\}$$

- Association
  - Points – points: feature matching

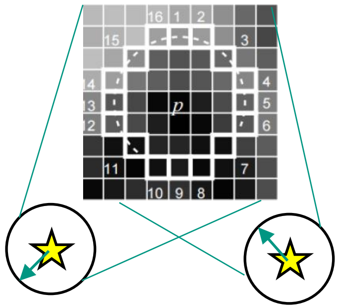
$$P_{asso} = |f(A) - f(B)| < thre\_f$$

  - Plane – plane: normal and distance

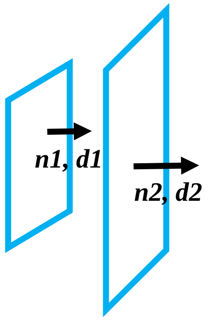
$$\pi_{asso} = (|n1 - n2| < thre\_n \; \&\& \; |d1 - d2| < thre\_d)$$

  - Object – object: 3D IoU

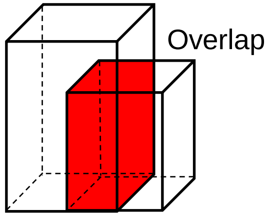
$$O_{asso} = \frac{V_{overlap}}{V_1 + V_2 - V_{overlap}} > thre\_v$$



Point-point association



Plane-plane association

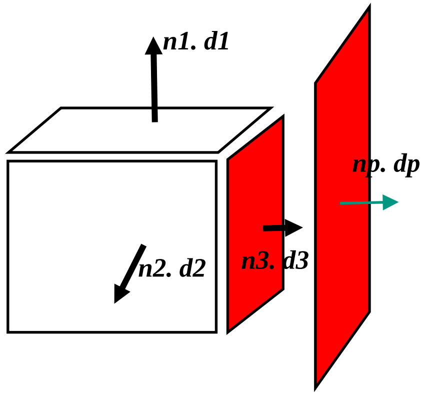


Object-object association

Method: Optimization

$$C^*, O^*, \Pi^*, P^* = \underset{\{C,O,\Pi,P\}}{\arg\min} \sum_{i \in C, j \in O, k \in \Pi, m \in P} \|e\|_\sigma^2$$

camera | objects | planes | points | constraints

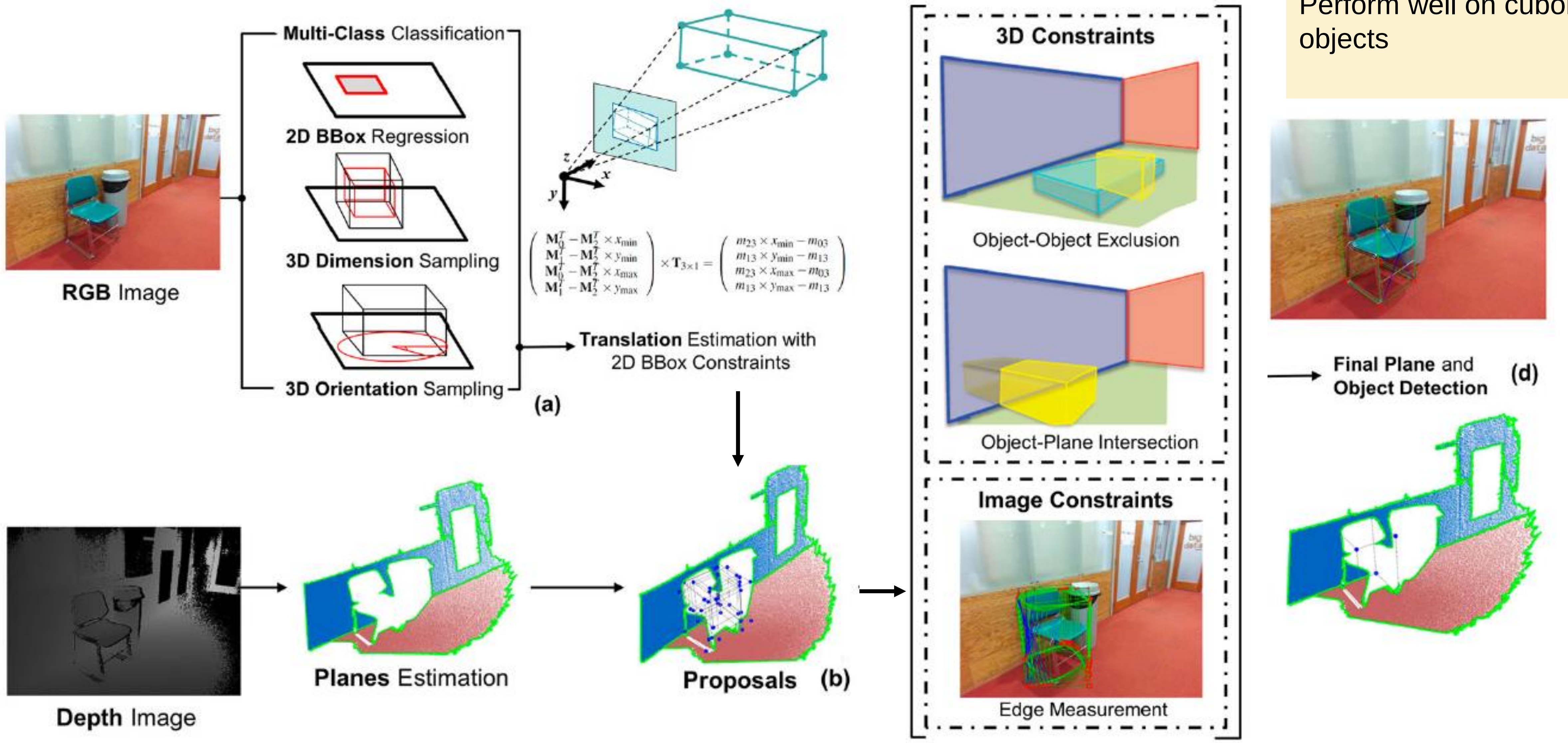


Plane-object constraints

- Constraints
  - Camera – points: geometric reprojection error $e(P_w, \mathbf{T}_{cw}) = u_c - \rho(\mathbf{T}_{cw}^{-1}, P_w)$ — $\rho(.)$ is to project 3D points to image
  - Camera – planes: angle and distance error $e(\pi_w, \mathbf{T}_{cw}) = \left\| q(\pi_m) - q\left(\mathbf{T}_{cw}^{-\top} \pi_w\right) \right\|$ — $q(\pi)$ is the minimum representation of plane
  - Camera – objects: corner reprejection error $e(O_w, \mathbf{T}_{cw}) = \sum_{m \in \{1,8\}} z_m - \rho(\mathbf{T}_{cw}^{-1}, O_w)$
  - Point – planes: orthogonal distance error $e(P, \Pi) = \|\pi P\|$ — πP is the orthogonal distance from point to plane
  - Point – objects: orthogonal distance error $e(P, O) = \max\left(\left|T_o^{-1} P\right| - \mathbf{d}_m, \mathbf{0}\right)$
  - Plane – objects: angle and distance error $e(\Pi, O) = \|\min(q(\pi) - q(\pi_{oi}))\|$

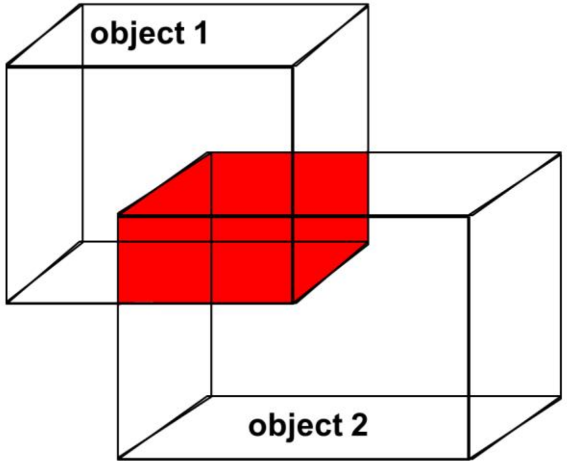
Method: Single Frame Object Detection

Mainly RGB information
Fast
Perform well on cuboid objects

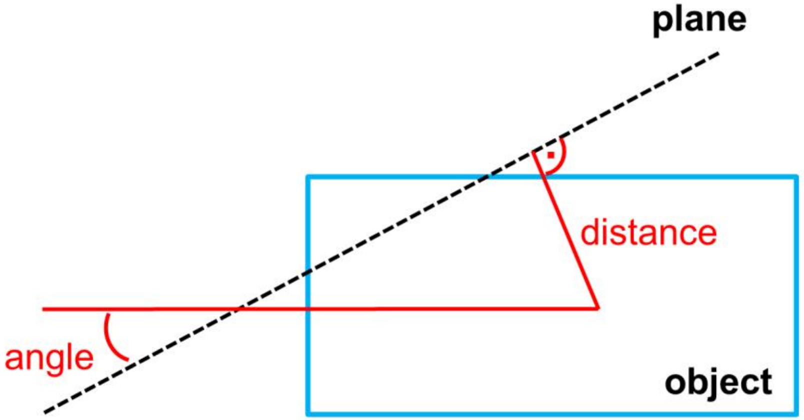


Method: Single Frame Object Detection

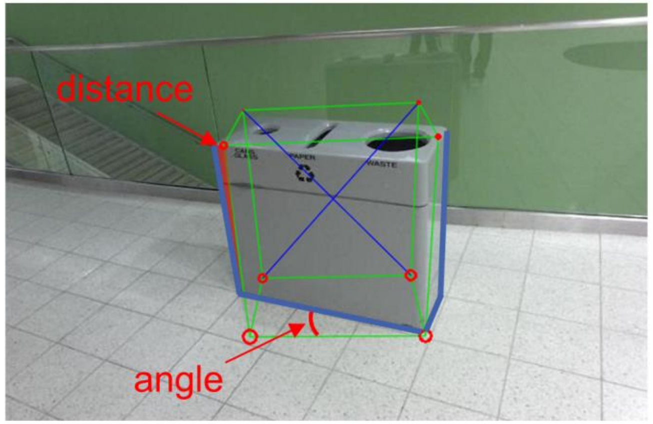
- Object selection with constraints



(a) Object-Object Constraints: 3D IoU
(*front view*)



(b) Object-Plane Constraints: angle & distance
(*top view*)



(c) Object-Image Constraints: angle & distance
(*projected view*)

Experiments: Single Frame Object Detection

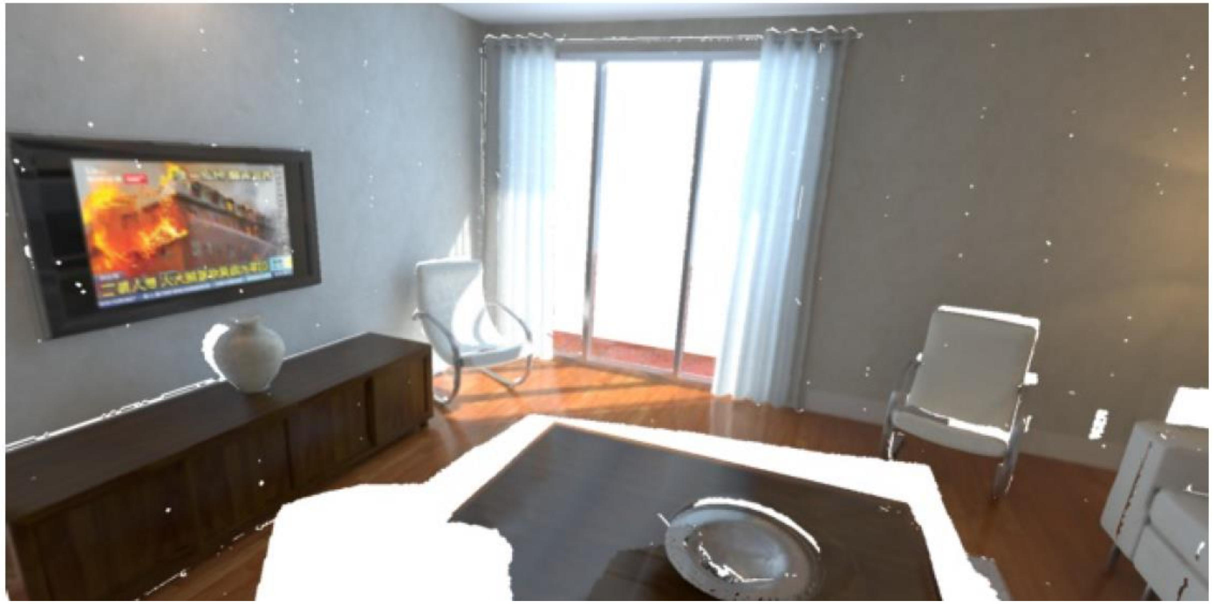
Experiments: Single Frame Object Detection

- SUN RGB-D Dataset

Class name (label_v6)	count	3D IoU (Intersection over Union)	Translation error (m)	Yaw error (rad)	Dimension error (m)
bed	184	0.4979	0.3963	0.1602=9.1°	[0.1527 0.1253 0.1126]
sofa	40	0.3189	0.4352	0.1569=9.0°	[0.1550 0.0968 0.0657]
sofa_chair	77	0.374	0.3244	0.1601=9.1°	[0.0469 0.0654 0.0575]
chair	163	0.3988	0.228	0.1637=9.4°	[0.0512 0.0477 0.0488]
garbage_bin	77	0.3786	0.1633	0.1090=6.2°	[0.0440 0.0391 0.0429]
night_stand	101	0.3559	0.2420	0.1634	[0.0505 0.0419 0.0537]
lamp	75	0.2578	0.2376	0.1371	[0.0368 0.0413 0.0462]
table	25	0.2576	0.5165	0.1804	[0.1227 0.0940 0.0787]

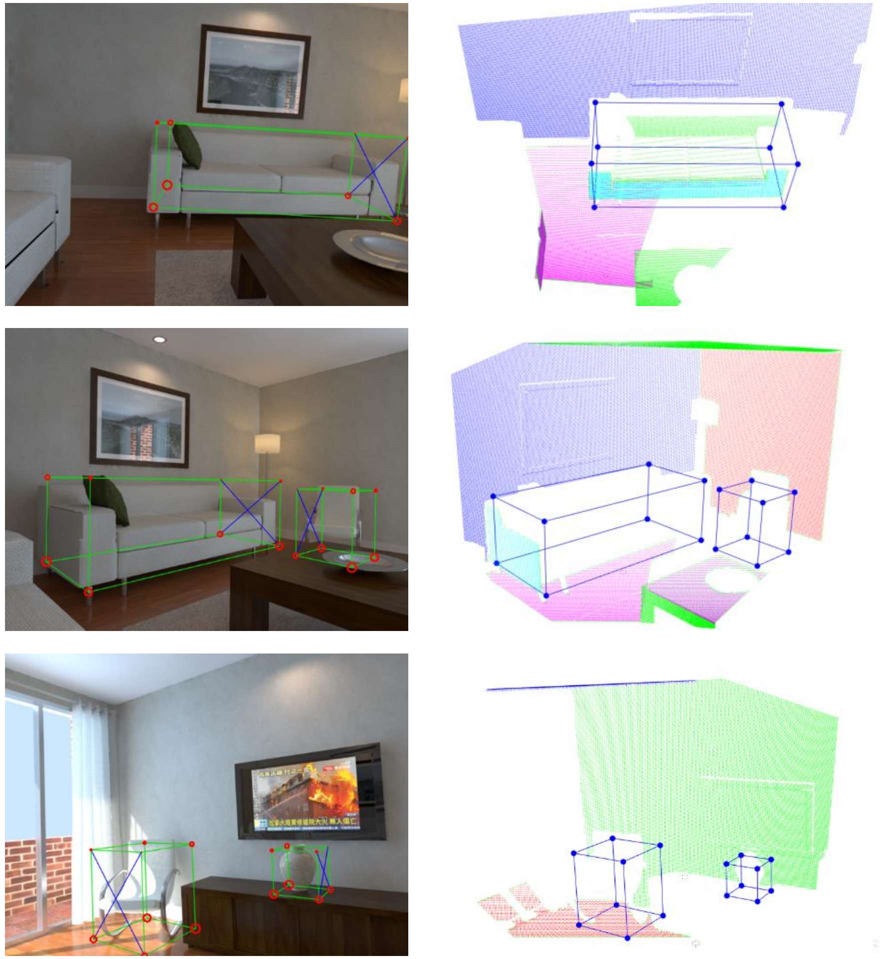
Experiments: SLAM

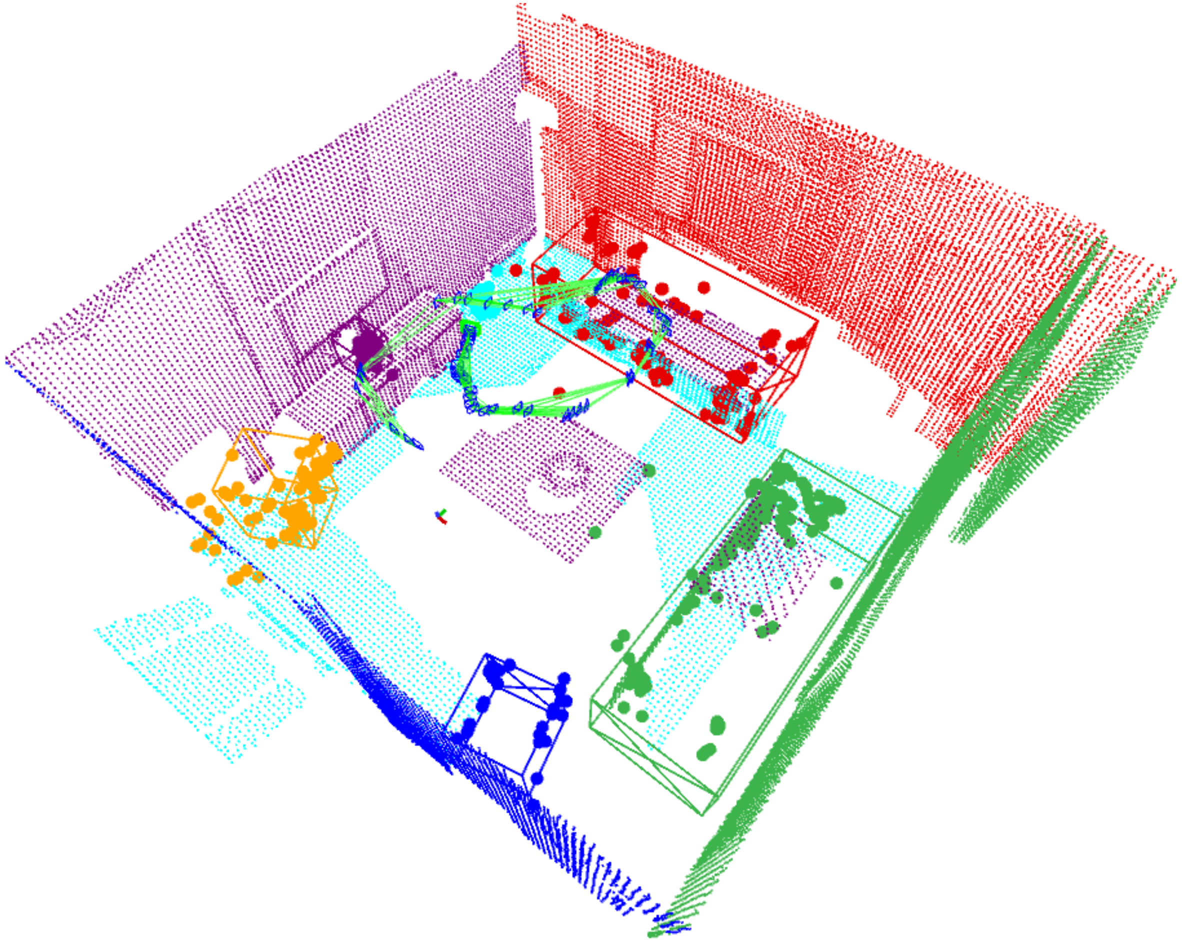
- Indoor dataset: ICL-NUIM (living room and office)
- Comparison
  - Points only SLAM (ORB SLAM 2)
  - Points + Plane SLAM
  - Points + Object SLAM
  - Points + Plane + Object SLAM (Ours)



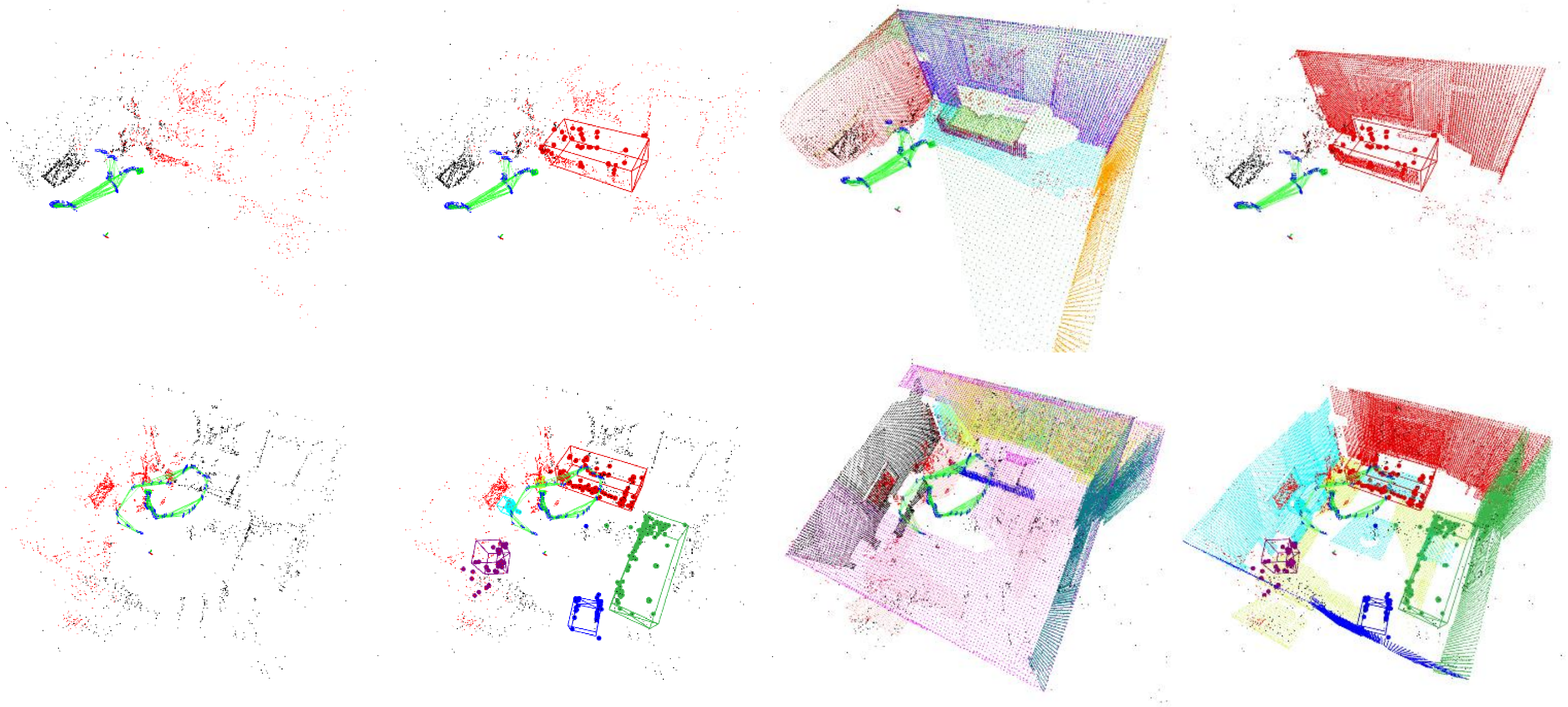
Experiments: SLAM

Karlsruher Institut für Technologie





	Point Map	Point Object Map	Point Plane Map	Point Plane Object Map
Living room kt-0				
Living room kt-2				

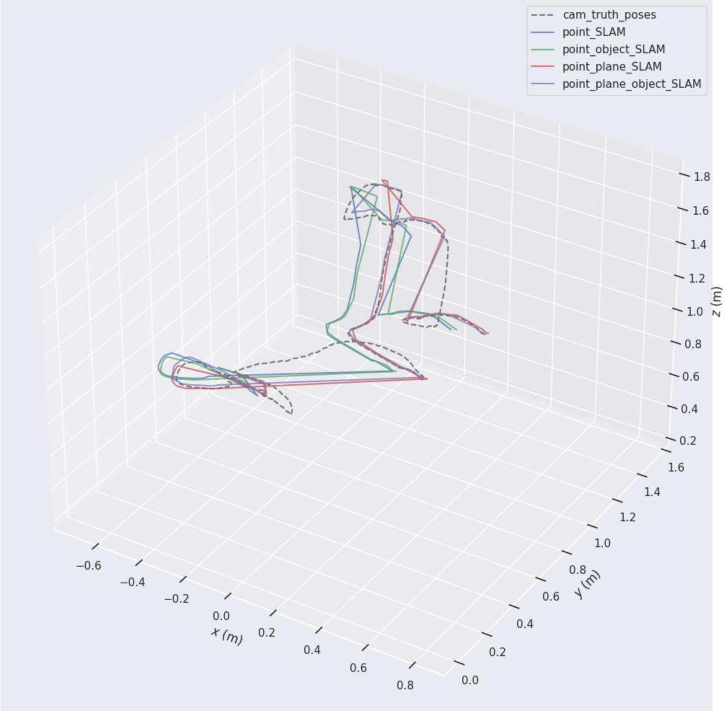


Experiments: SLAM

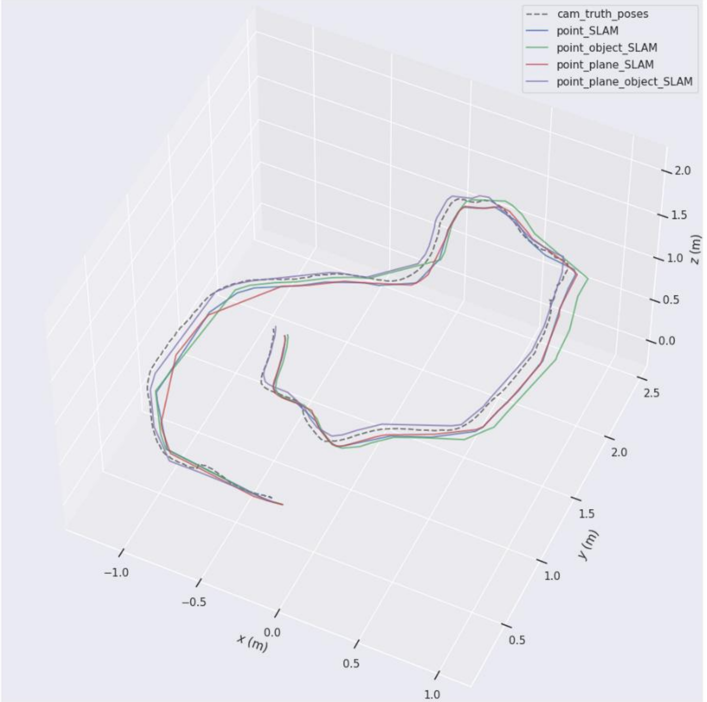
- Table: Evaluation results of root mean squared error of absolute camera pose error (RMSE-APE) on dataset (cm)
- Figure: Comparison of the estimated trajectories and corresponding ground truth.

Method	Living room kt-0	Living room kt-2	Living room kt-3
Point only SLAM	**0.3971**	2.5773	2.6434
Point-Plane SLAM	0.4927	2.3053	2.5570
Point-Object SLAM	0.4021	2.0543	2.5520
Point-Plane-Object SLAM	0.9916	**1.77044**	**1.7551**

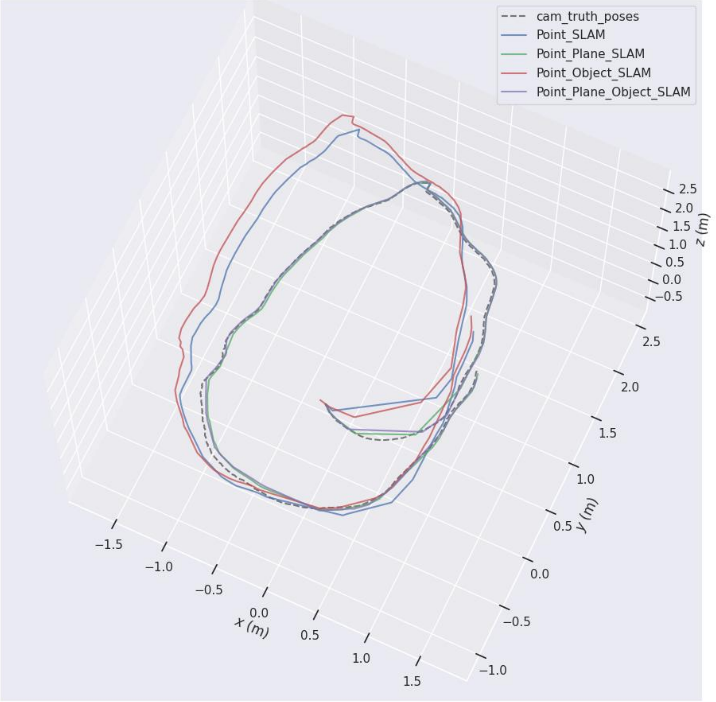
--- GT Camera — Point SLAM — Point-Plane SLAM — Point-Object SLAM — Point-Plane-Object SLAM



a) Living room kt-0

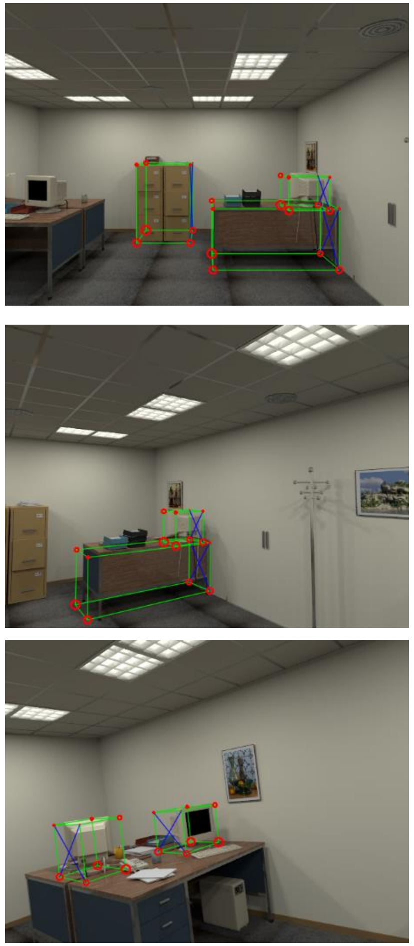


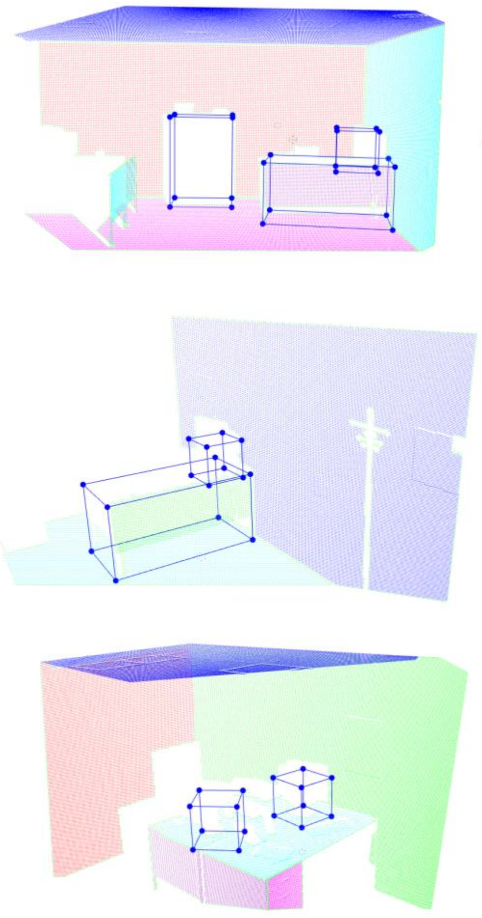
b) Living room kt-2

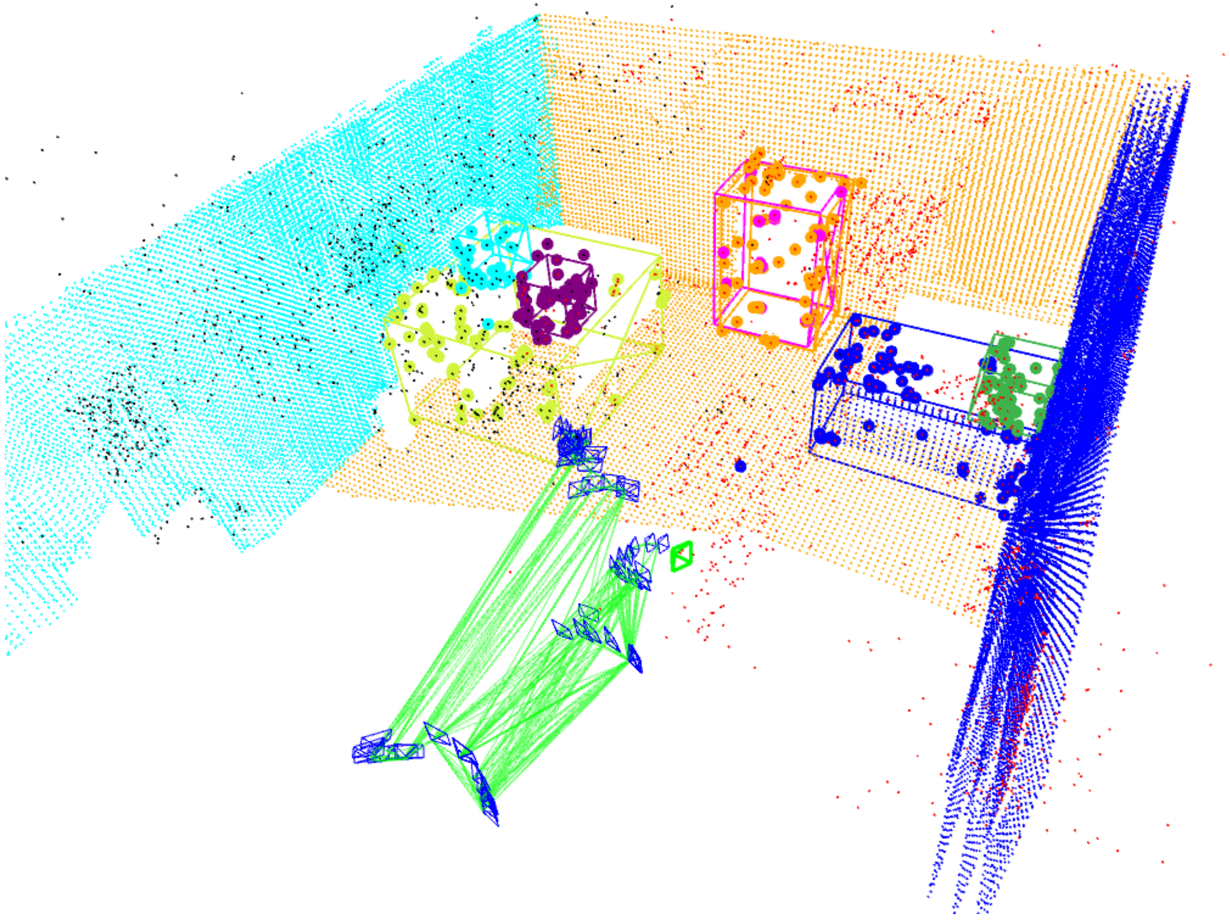


c) Living room kt-3

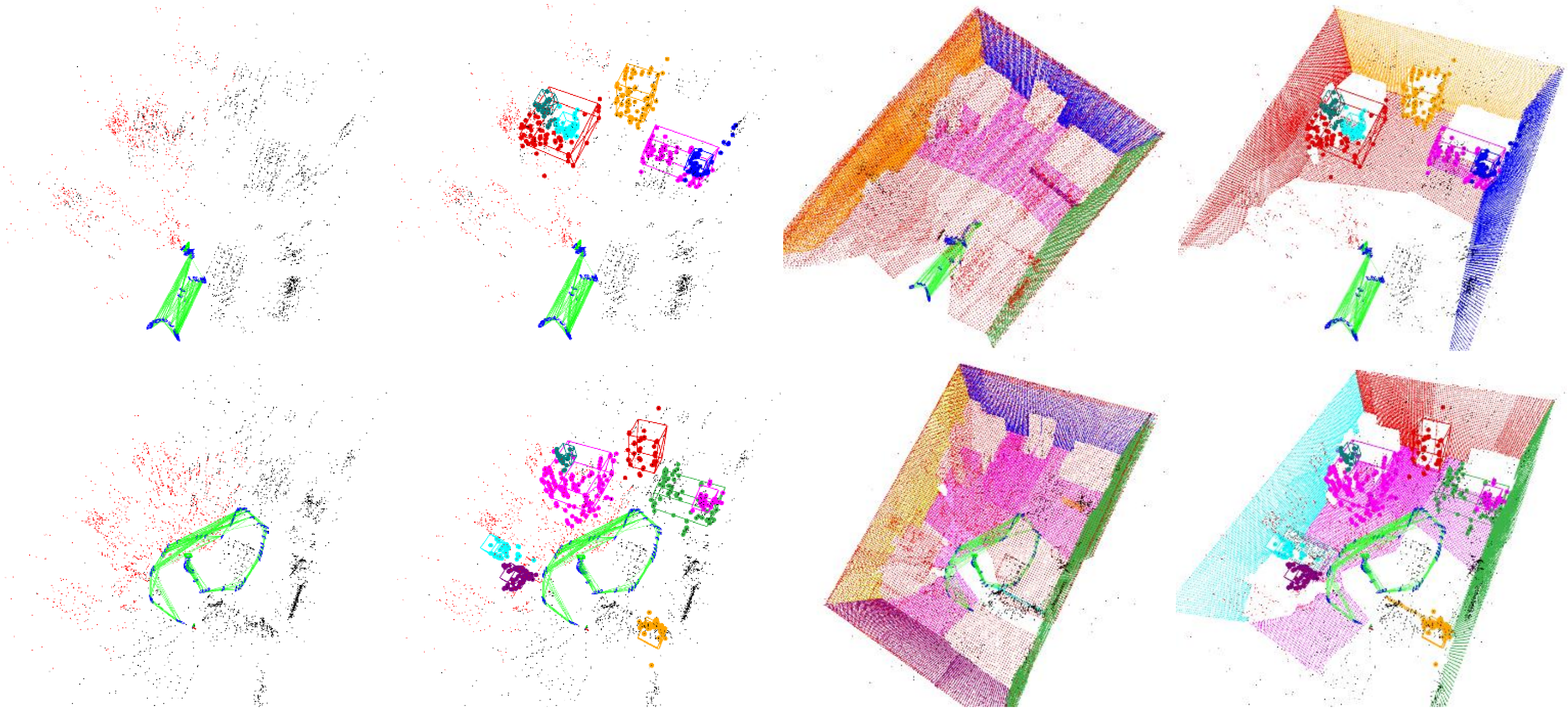
Experiments: SLAM







	Point Map	Point Object Map	Point Plane Map	Point Plane Object Map
Office kt-0				
Office kt-2				

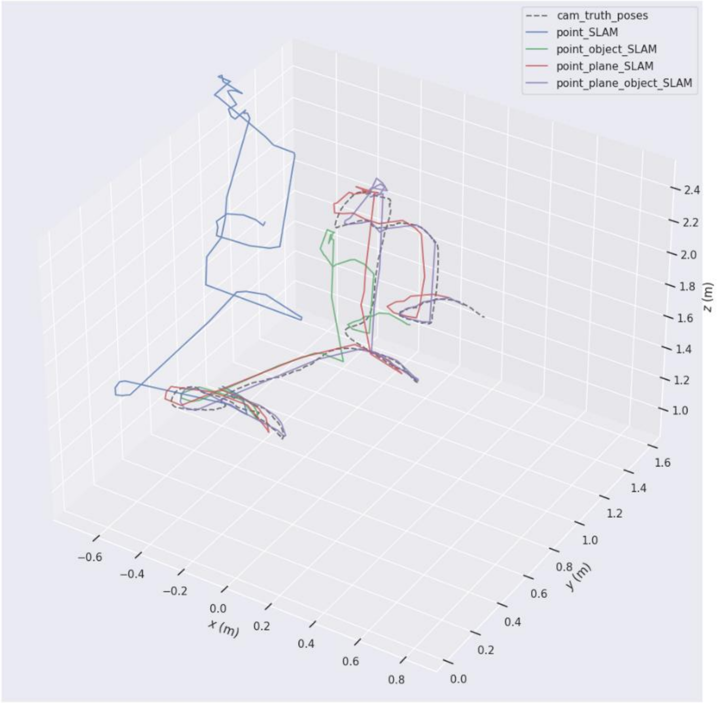


Experiments: SLAM

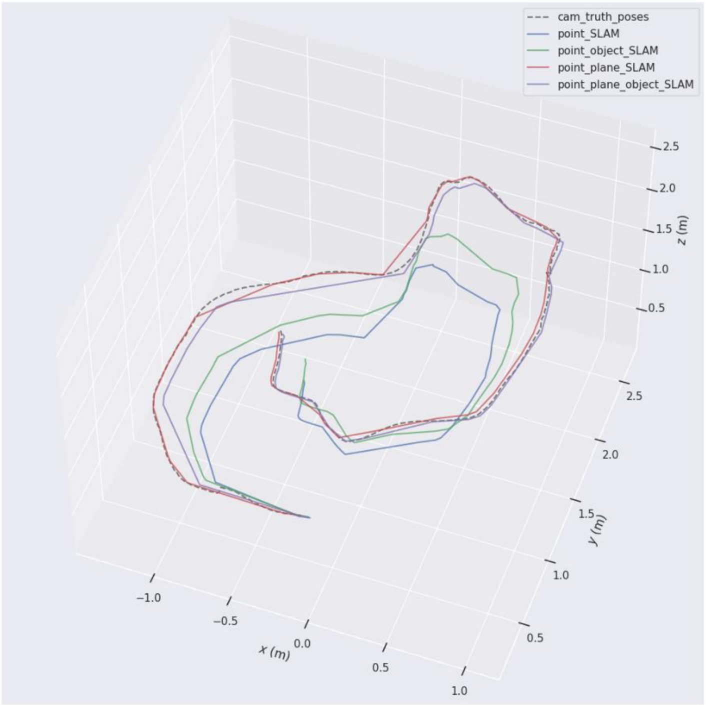
- Table: Evaluation results of root mean squared error of absolute camera pose error (RMSE-APE) on dataset (cm)
- Figure: Comparison of the estimated trajectories and corresponding ground truth.

Method	Office kt-0	Office kt-2	Office kt-3
Point only SLAM	12.8362	2.6126	3.9202
Point-Plane SLAM	**5.8526**	**1.4260**	**2.9024**
Point-Object SLAM	6.3235	5.3889	3.1101
Point-Plane-Object SLAM	6.8302	1.7747	3.66788

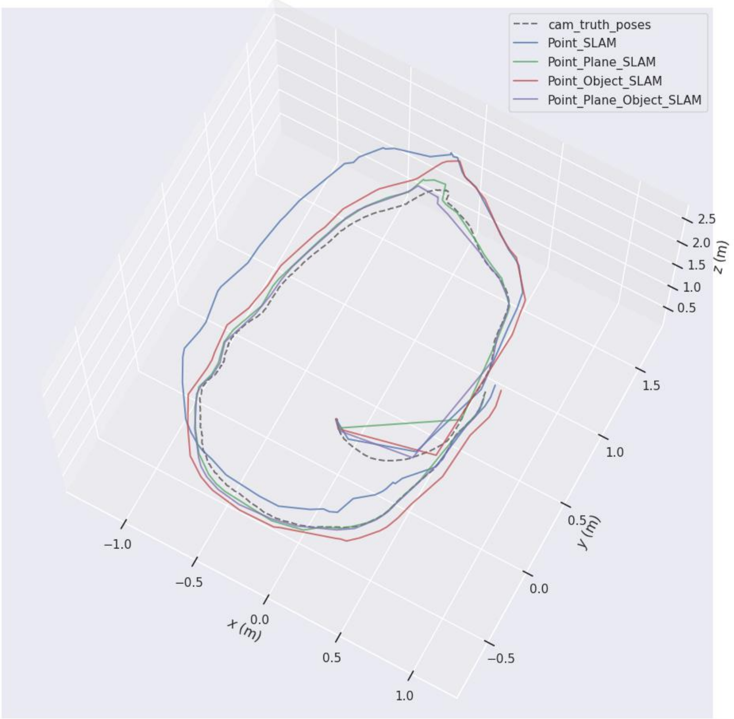
--- GT Camera — Point SLAM — Point-Plane SLAM — Point-Object SLAM — Point-Plane-Object SLAM



a) Office kt-0



b) Office kt-2



c) Office kt-3

Experiments: SLAM

- Table: Average runtime of different SLAM components in ICL NUIM living room kt-2 sequence.

	Tasks	Average time (mSec)
Single image preprocess	Plane estimation	109.99
	Object detection	97.386
	Edge detection	18.831
Indoor ICL room dataset	Tracking thread	47.886
	Point only BA	63.240
	Point plane BA	135.55
	Point plane object BA	157.48

Conclusion

- What we have done
  - Proposed a structure SLAM which adds planes and objects to existing ORB-SLAM 2 system realize camera localization
  - Evaluated proposed SLAM method in and office sequences, results showed the introduction of objects can slightly improve localization accuracy