Real-Time Monocular Visual SLAM by Combining Points and Lines

Xinyu Wei, Zengming Tang, Huiyan Wu, Jun Huang*

Shanghai Advanced Research Institute, Chinese Academy of Sciences

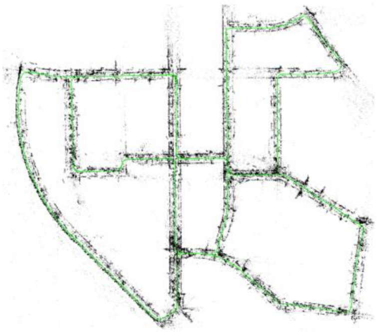
Outline

- Background
- SLAM Combining Points and Lines
- Line Features in Visual SLAM
- Results
- Conclusion

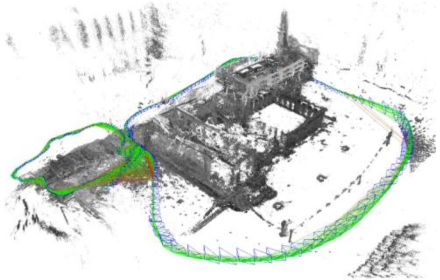
Background

What is Visual SLAM?

- **SLAM**: **S**imultaneous **L**ocalization **A**nd **M**apping.
- **Mapping**: Constructing a map of an unknown environment.
- **Localization**: Estimating the six degrees of freedom (DoF) robot motion.



ORB SLAM, 2015
R Mur-Artal et al.

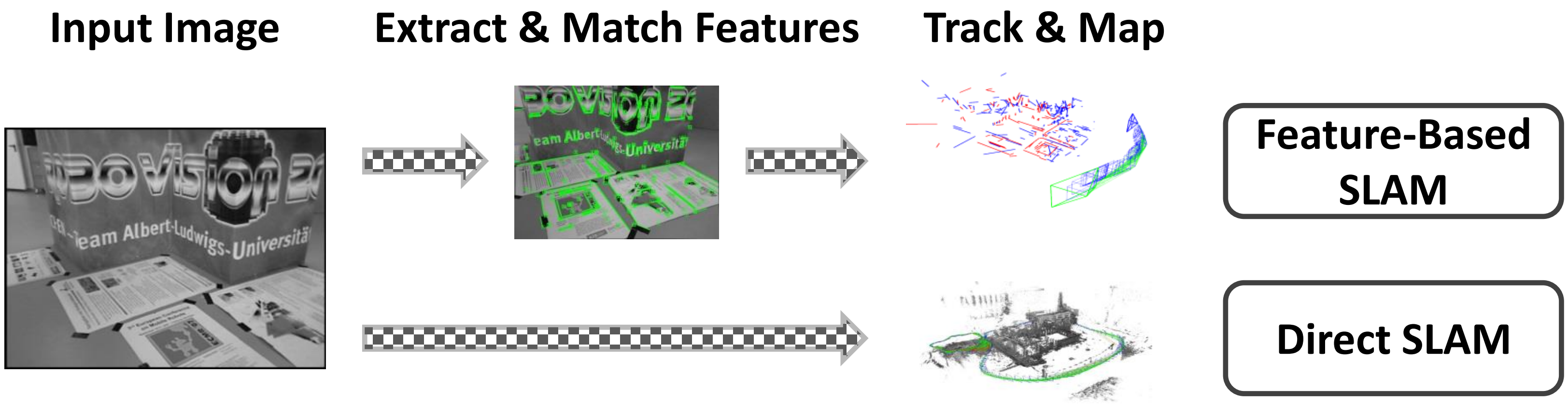


LSD-SLAM, 2014
J Engel et al.

Robotics
Augmented Reality (AR)

Background

Direct SLAM & Feature-Based SLAM



- **Information**: Full image vs. Features (ORB, SIFT, LSD, …)
- **Tracking**: Minimizing photometric error vs. Minimizing reprojection error
- **Mapping**: Per-pixel depth vs. 3D points/lines/…

Background

Direct SLAM & Feature-Based SLAM

	Absolute KeyFrame Trajectory RMSE (cm)			
	ORB-SLAM	PTAM	LSD-SLAM	RGBD- SLAM
fr1_xyz	**0.90**	1.15	9.00	1.34 (1.34)
fr2_xyz	0.30	**0.20**	2.15	2.61 (1.42)
fr1_floor	**2.99**	X	38.07	3.51 (3.51)
fr1_desk	**1.69**	X	10.65	2.58 (2.52)
fr2_360 _kidnap	3.81	**2.63**	X	393.3 (100.5)
fr2_desk	**0.88**	X	4.57	9.50 (3.94)
fr3_long _office	**3.45**	X	38.53	–
fr3_nstr_tex_far	ambiguity detected	4.92 / 34.74	18.31	–
fr3_nstr_ tex_near	**1.39**	2.74	7.54	–
fr3_str_tex_far	**0.77**	0.93	7.95	–
fr3_str_ tex_near	1.58	**1.04**	X	–
fr2_desk_person	**0.63**	X	31.73	6.97 (2.00)
fr3_sit_xyz	**0.79**	0.83	7.73	–
fr3_sit_halfsph	**1.34**	X	5.87	–
fr3_walk_xyz	**1.24**	X	12.44	–
fr3_walk_halfsph	**1.74**	X	X	–

Motivation

More information is needed to improve the robustness and accuracy of feature point based SLAM system.

In the direct methods, the benefits of robustness and invariance to photometric variations which are provided by features are sacrificed.

Lines are usually abundant in manmade scenes.

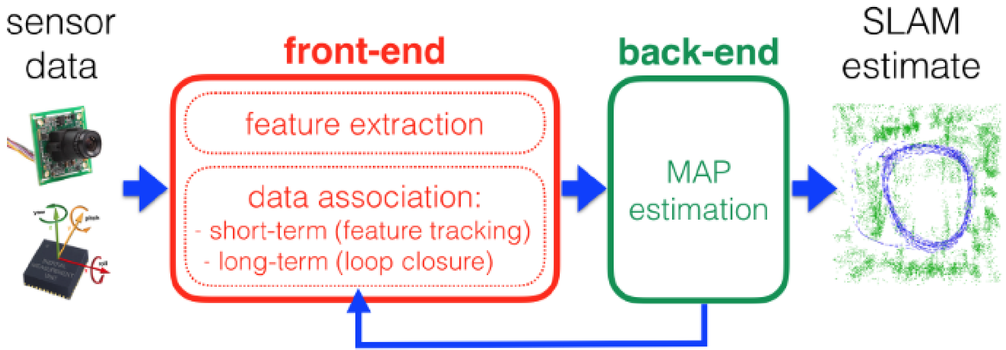
(The experimental results are taken from ORB SLAM, 2015, R Mur-Artal et al.)

Outline

- Background
- SLAM Combining Points and Lines
- Line Features in Visual SLAM
- Results
- Conclusion

Typical SLAM System

- **The front-end**: Abstracting sensor data into models that are amenable for estimation.
- **The back-end**: Performing inference on the abstracted data produced by the front-end.



(The image is taken from SLAM: Present, Future, and the Robust-Perception Age, 2015, C Cadena et al.)

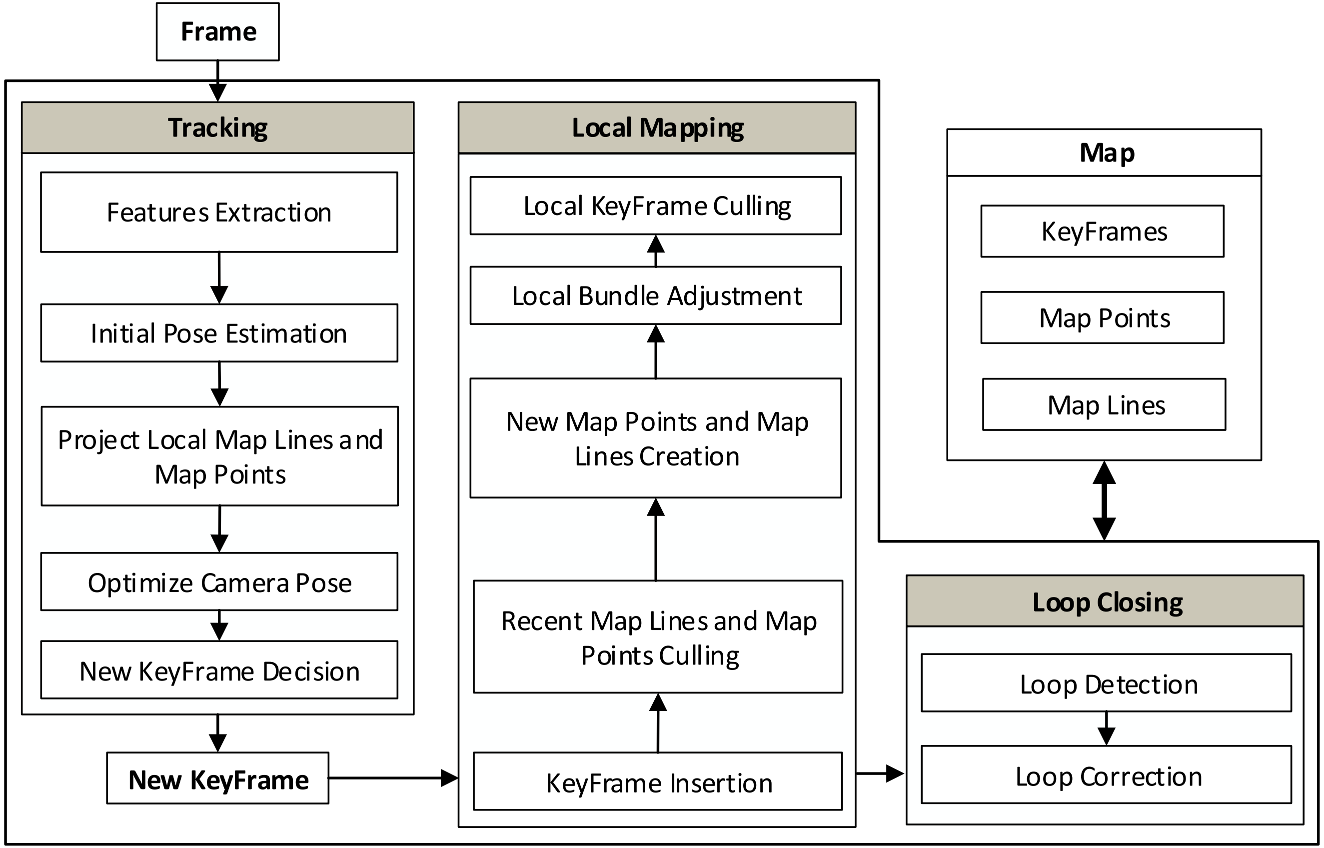
SLAM Combining Points and Lines

The front-end: Tracking; **The back-end**: Local Mapping, Loop Closing.

Tracking: Localizing camera pose in every frame and selecting new keyframe.

Local Mapping: Maintaining the dynamic map which includes keyframes, map points (3D points in map) and map lines (3D line segments in map).

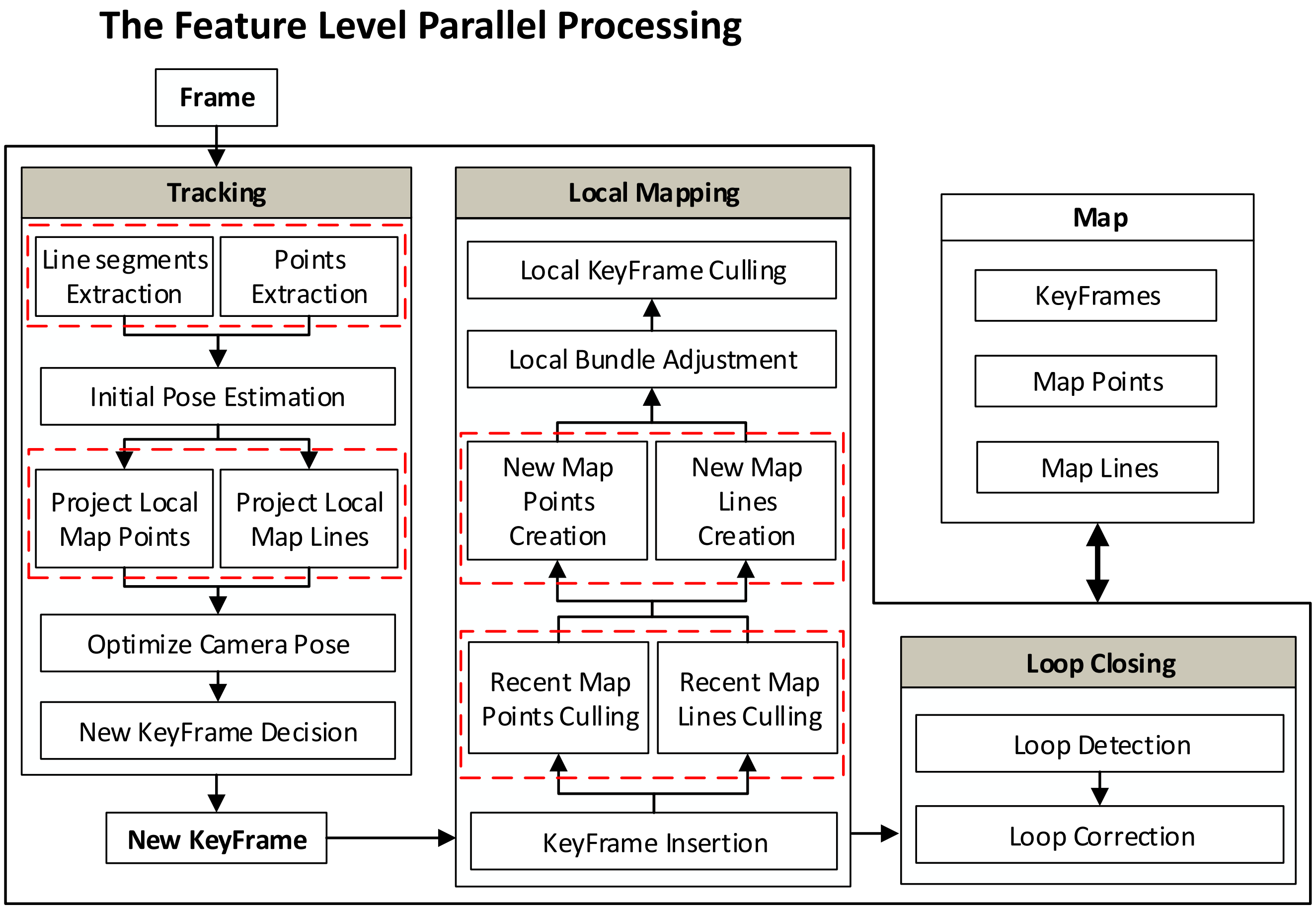
Loop Closing: Searching loop and corrects it.



SLAM Combining Points and Lines

Problem:
High computational burden is required for the multi-feature processing tasks. In the existing similar SLAM systems, only a few barely reached the real-time specifications (20 Hz).

The proposed system can run at around 30 Hz.



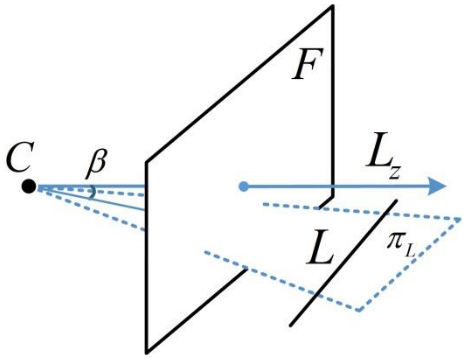
Outline

- Background
- SLAM Combining Points and Lines
- Line Features in Visual SLAM
- Results
- Conclusion

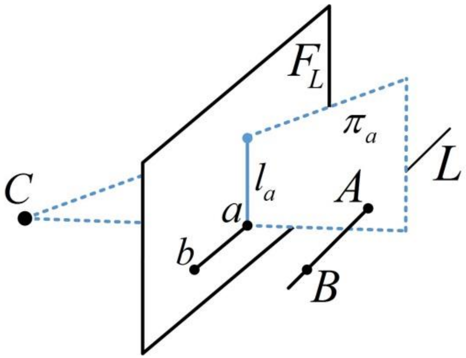
Line Features in Visual SLAM

Map Lines in System

- **Map Lines**: $L = (L_P, L_F)$
- L_P: The 3D line containing the line segment which is parameterized with Plücker line coordinate and orthonormal representation;
- L_F: A set of keyframes in which the line can be observed.



Observation angle β of L in the keyframe F

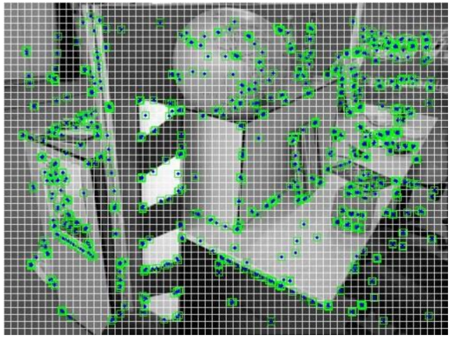


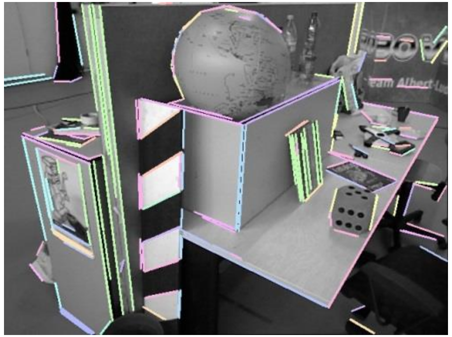
Calculation of the endpoints A, B of L

Line Features in Visual SLAM

A Fast Line Matching Algorithm

- **Task**: Finding the best matching line l of map line L in frame F.
  1. Projecting the endpoints A and B of L in F, getting the projection line segment l_L.
  2. Computing the candidate matching line segments of l_L in F.
  3. Calculating the best matching line of L.





$$\begin{aligned} d_1(l_L, l) &= \alpha(d(l_L, p_1) + d(l_L, p_2)) \\ d_2(l_L, l) &= (1-\alpha)(d'(p_1, a) + d'(p_2, b)) \\ d(l_L, l) &= (d_1(l_L, l) + d_2(l_L, l))\frac{l_F^*}{l_L^*} \end{aligned}$$

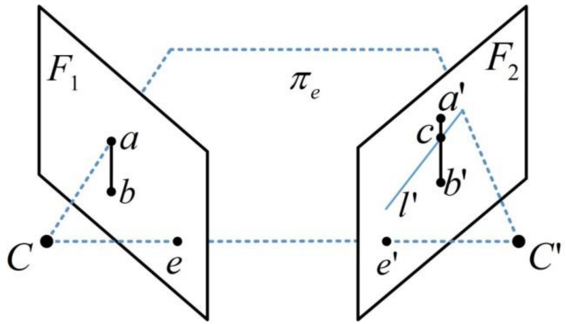
Classification of the features

Matching error of segments l_L and l

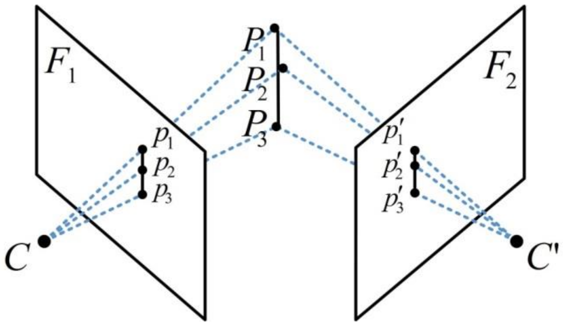
Line Features in Visual SLAM

Map Line Initialization

- **Task**: Using line segment matches and camera poses to calculate new map lines.
  1. Obtaining sampling point matches on 2D lines in two frame.
  2. Calculating 3D points according to sampling point matches.
  3. Creating map lines with line fitting.



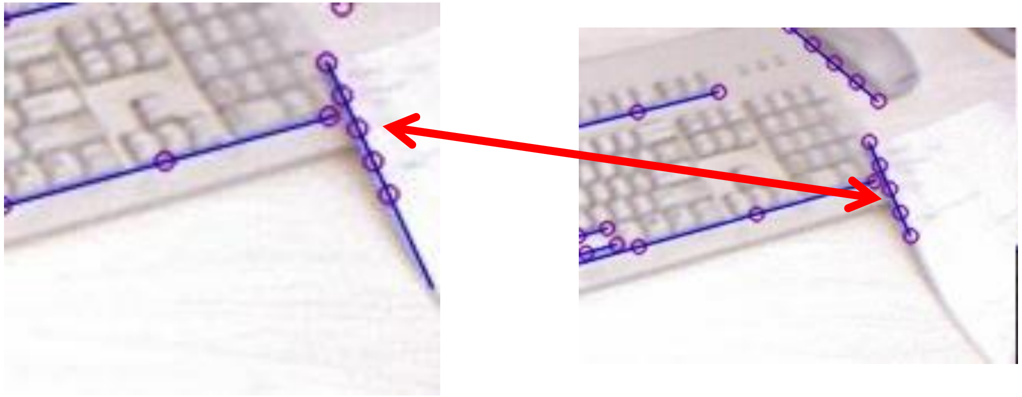
Sampling points matching based on epipolar constraint



Map line fitting

Line Features in Visual SLAM

Map Line Initialization



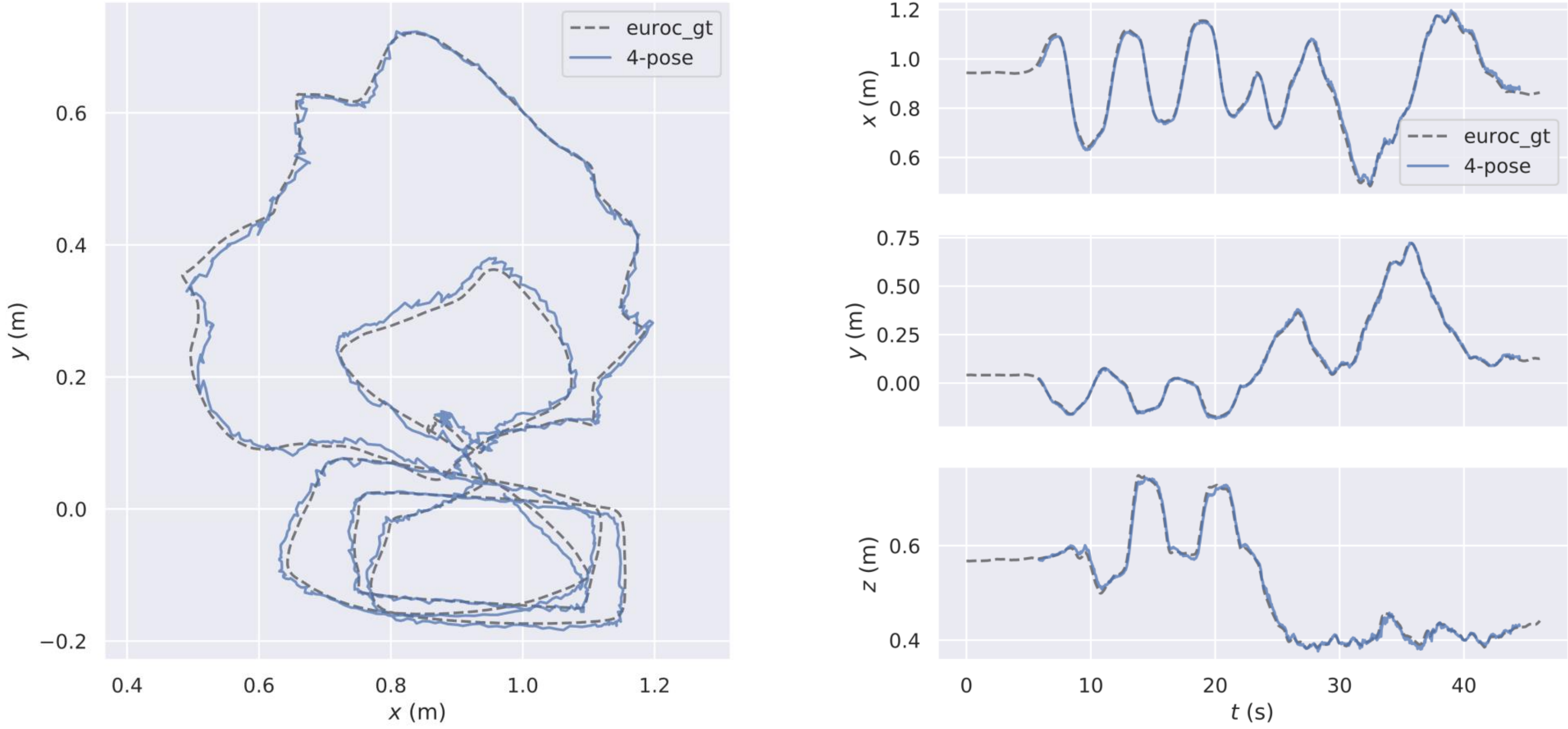
Accurate sampling point matching with unreliable line endpoints

Outline

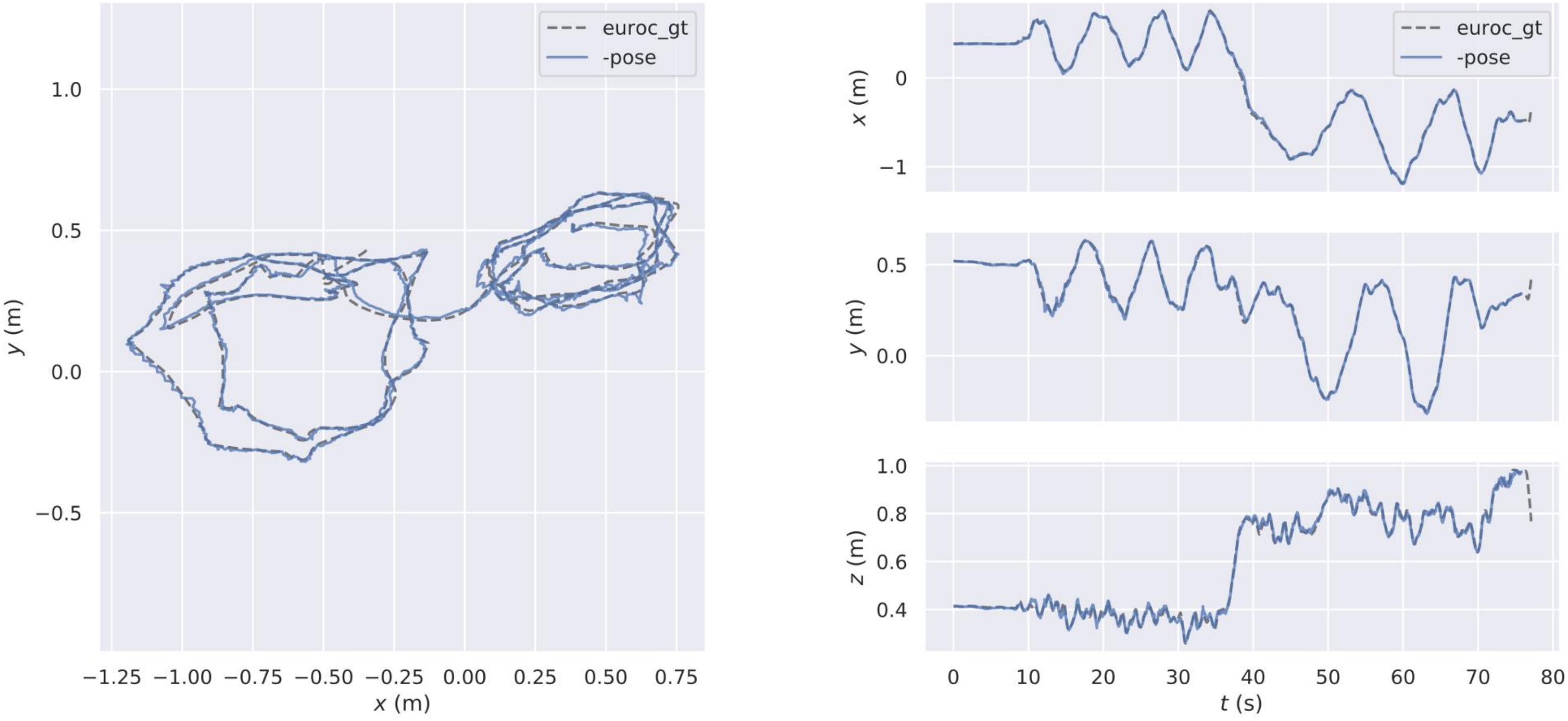
- Background
- SLAM Combining Points and Lines
- Line Features in Visual SLAM
- Results
- Conclusion

Results - Localization Accuracy on C3_train

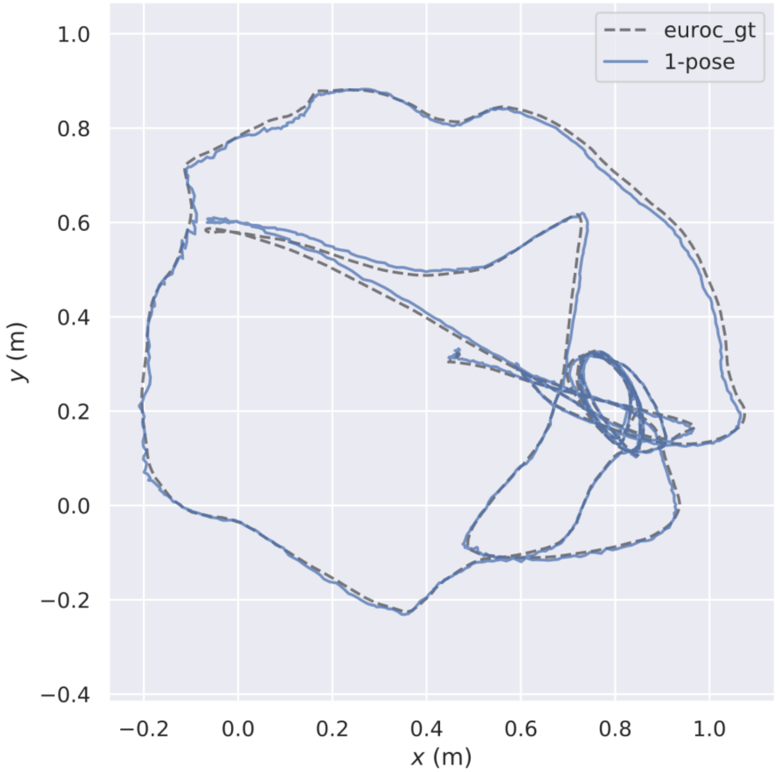
- Intel Core i7-3770 (4 cores @3.4GHz) , 4 GB RAM, No GPU

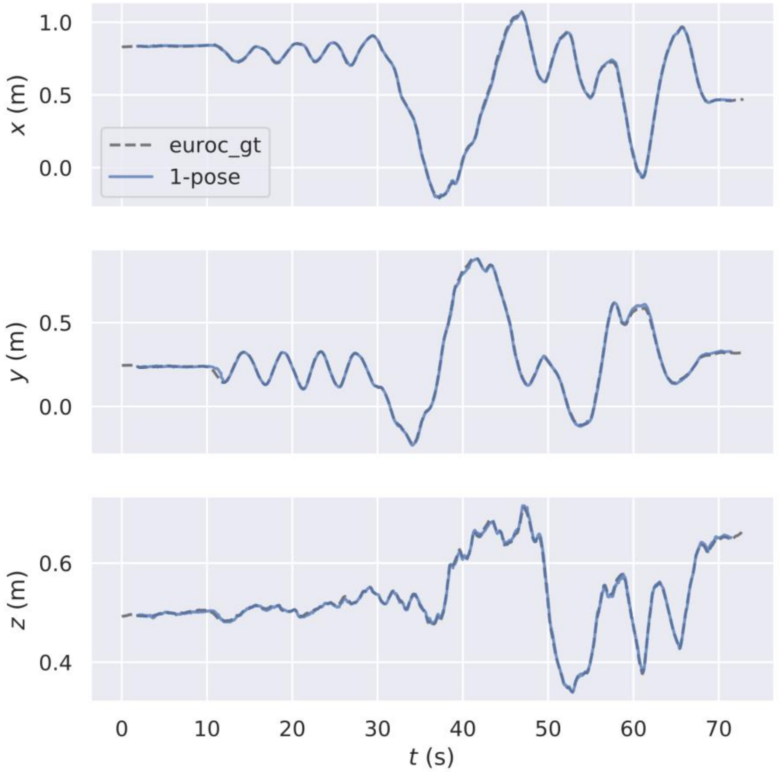


Results - Localization Accuracy on C5_train

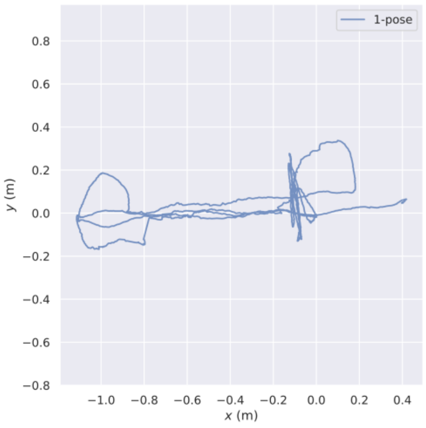


Results - Localization Accuracy on C6_train

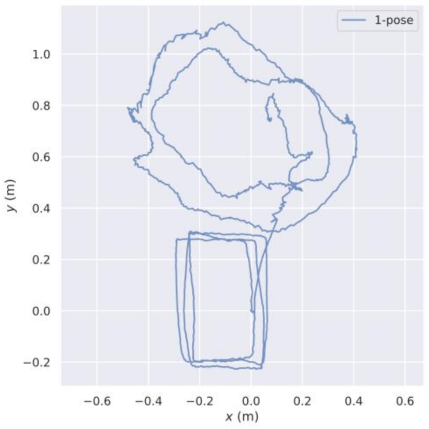




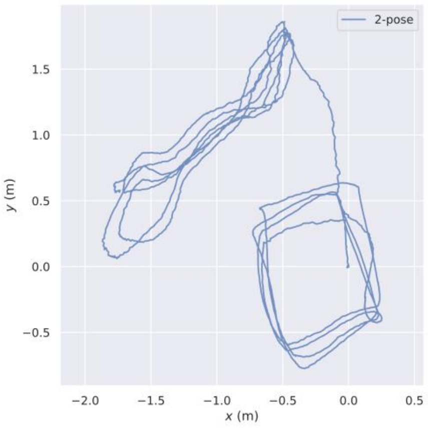
Results - Camera trajectories on The Test Dataset



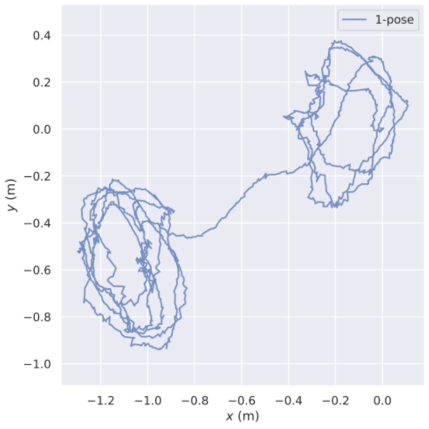
C2_test



C3_test



C4_test



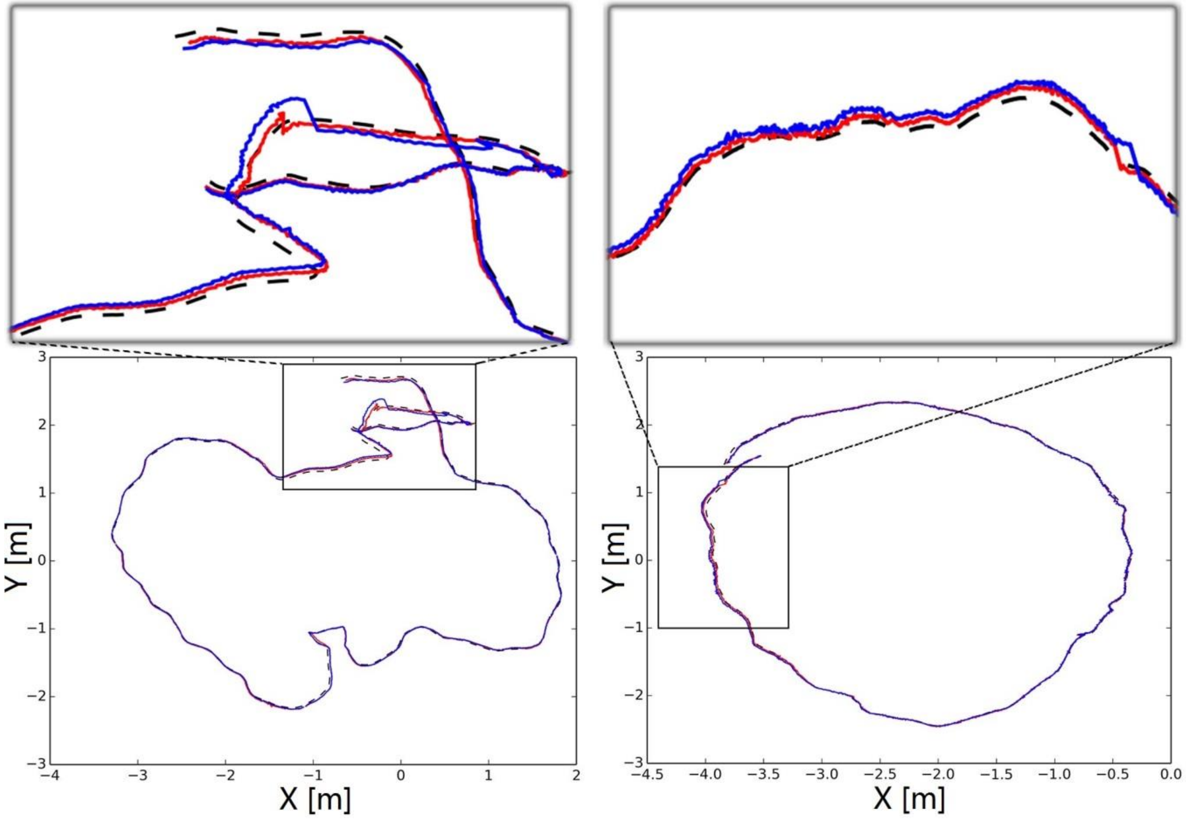
C5_test

Results - Localization Accuracy on The Training Dataset

Sequence		PTAM		ORB-SLAM2		DSO		PL-SLAM	
APE/RPE (mm)	C0	**75.442**	6.696	96.777	**5.965**	231.86	10.456	82.780	7.028
	C1	113.406	16.344	95.379	**10.285**	431.929	12.555	**68.875**	22.223
	C2	67.099	6.833	69.486	5.706	216.893	**5.337**	**51.400**	9.970
	C3	**10.913**	4.627	15.310	7.386	188.989	**4.294**	12.259	6.506
	C4	21.007	4.773	**10.061**	**2.995**	115.477	4.595	30.015	4.437
	C5	40.403	8.926	29.653	11.717	323.482	**7.978**	**18.973**	8.418
	C6	19.483	**3.051**	12.145	6.741	14.864	2.561	**9.435**	3.080
	C7	13.503	2.462	**5.832**	**1.557**	27.142	2.213	15.413	3.225
ARE/RRE (deg)	C0	12.051	**0.257**	5.119	0.342	9.983	0.401	**4.605**	0.436
	C1	53.954	0.291	**8.534**	0.242	39.007	0.524	33.426	**0.193**
	C2	8.789	0.301	5.550	**0.255**	10.584	0.253	**2.640**	0.577
	C3	6.225	0.293	1.431	0.264	20.580	**0.241**	**1.287**	0.624
	C4	6.295	0.255	**1.015**	**0.157**	5.217	0.180	2.624	0.221
	C5	14.03	0.452	1.963	0.546	40.939	**0.324**	**1.933**	0.563
	C6	2.348	0.217	**0.892**	**0.169**	1.435	0.189	1.068	0.304
	C7	1.218	0.153	**0.569**	**0.115**	2.239	0.135	0.779	0.211
Completeness (%)	C0	79.386		65.175		14.476		**79.696**	
	C1	60.893		**68.303**		0.869		17.131	
	C2	**85.348**		79.263		22.878		81.437	
	C3	71.635		98.497		43.493		**100.000**	
	C4	95.418		**100.000**		80.371		99.895	
	C5	87.399		97.785		2.059		**100.000**	
	C6	97.399		99.786		**100.000**		**100.000**	
	C7	**100.000**		**100.000**		**100.000**		99.122	

(The experimental results of PTAM, ORB-SLAM2 and DSO are taken from VI-SLAM Survey for AR, 2019, Jinyu LI et al.)

Results - Localization Accuracy on The TUM RGB-D Dataset



Results - Localization Accuracy on The TUM RGB-D Dataset

TUM RGB-D Sequence	Absolute KeyFrame Trajectory RMSE (cm)					
	Ours	PL-SLAM	ORB-SLAM	PTAM	LSD-SLAM	RGBD-SLAM
f1_xy	**0.87**	1.21	0.94	1.15	9	1.34
f2_xyz	0.25	0.43	0.23	**0.2**	2.15	1.42
f3_long_office	**1.10**	1.97	1.68	-	38.53	-
f3_nstr_tex_near	**1.40**	1.58	1.43	2.74	7.54	-
f3_str_tex_far	0.97	**0.89**	1.05	0.93	7.95	-
f3_str_tex_near	1.16	1.25	1.19	**1.04**	-	-
f2_desk_person	**0.64**	1.99	0.72	-	31.73	2
f3_sit_xyz	0.81	**0.066**	0.89	0.83	7.73	-
f3_sit_halfsph	1.56	**1.31**	1.40	-	5.87	-
f3_walk_xyz	**1.17**	1.54	1.56	-	12.44	-
f3_walk_halfsph	1.68	**1.6**	1.93	-	-	-

(The experimental results of PTAM, ORB-SLAM2 and DSO are taken from VI-SLAM Survey for AR, 2019, Jinyu LI et al.)

Results

Running Time of Each Operation (TUM DataSet)

Thread	Operation	Mean execution time (ms)			
		Ours-A	Ours-B	PL-SLAM	ORB-SLAM
Tracking	Features extraction	14.94	24.37	31.32	12.36
	Initial Pose Estimation	4.82	4.75	7.16	3.27
	Track Local Map	9.22	9.41	12.58	6.93
	Total	28.98	38.53	51.06	22.56

Thread	Operation	Mean execution time (ms)			
		Ours-A	Ours-B	PL-SLAM	ORB-SLAM
Local Mapping	KeyFrame Insertion	15.38	19.18	17.08	13.99
	Map Feature Culling	0.44	0.49	1.18	0.08
	Map Feature Creation	190.94	231.22	74.64	76.74
	Local BA	154.06	179.68	218.25	107.95
	KeyFrame Culling	2.94	3.78	12.7	6.27
	Total	363.76	434.35	323.85	205.03

Outline

- Background
- SLAM by Combining Points and Lines
- Line Features in Visual SLAM
- Results
- Conclusion

Conclusion

- A real time monocular SLAM system using both points and line segments.
- Novel line processing algorithms in visual SLAM.
- A feature level parallel processing framework.

Future Work

- Using geometric constraints (coplanar, vertical, et al.) of features to improve the reconstruction accuracy of map.
- Multi-sensor (camera, IMU, GPS, et al.) information fusion in SLAM.

Thanks!