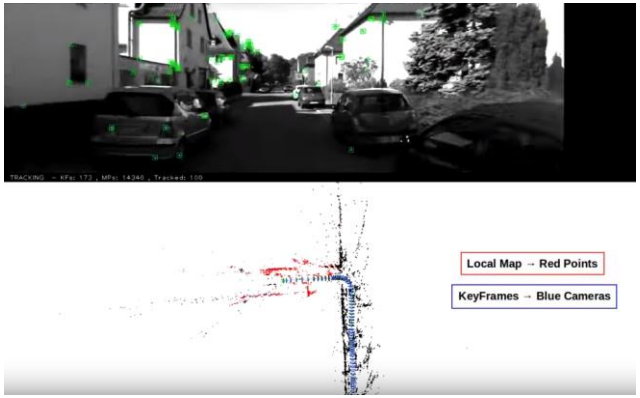
ORB-SLAM

A Versatile and Accurate Monocular SLAM System

Master-Seminar:
The Evolution of Motion Estimation and Real-time 3D Reconstruction

Julia Kabalar, 04 Dezember 2019



Outline

- Introduction
- Method Description
- Experiments & Results
- Resume

Outline

- Introduction
- Method Description
- Experiments & Results
- Resume

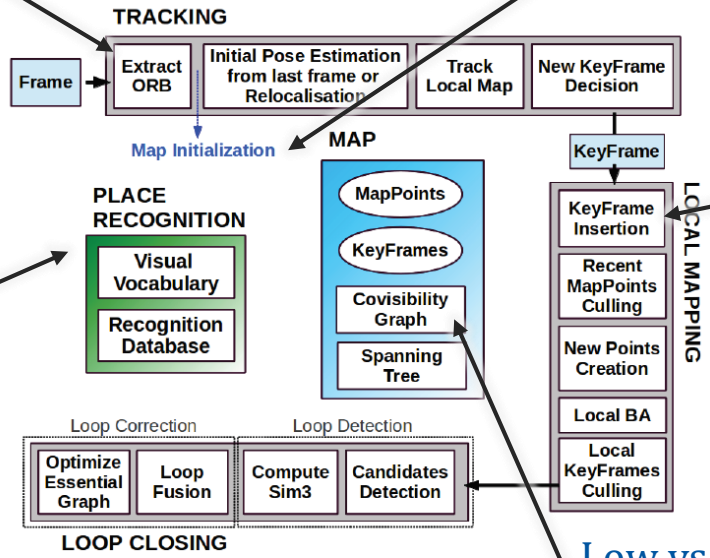
Introduction

- **Monocular** SLAM
- **Indirect** method
- **Sparse** map
- uses co-visibility information
- keyframes
- pose graph-based optimization
- with loop-closure

PTAM versus ORB-SLAM

search by patch correlation vs. ORB descriptor

stereo-pair vs. automatic initialization



Cautious insertion vs. survival of the fittest

New module works with ORB features

Low vs. fast convergence

Map Components

Map points

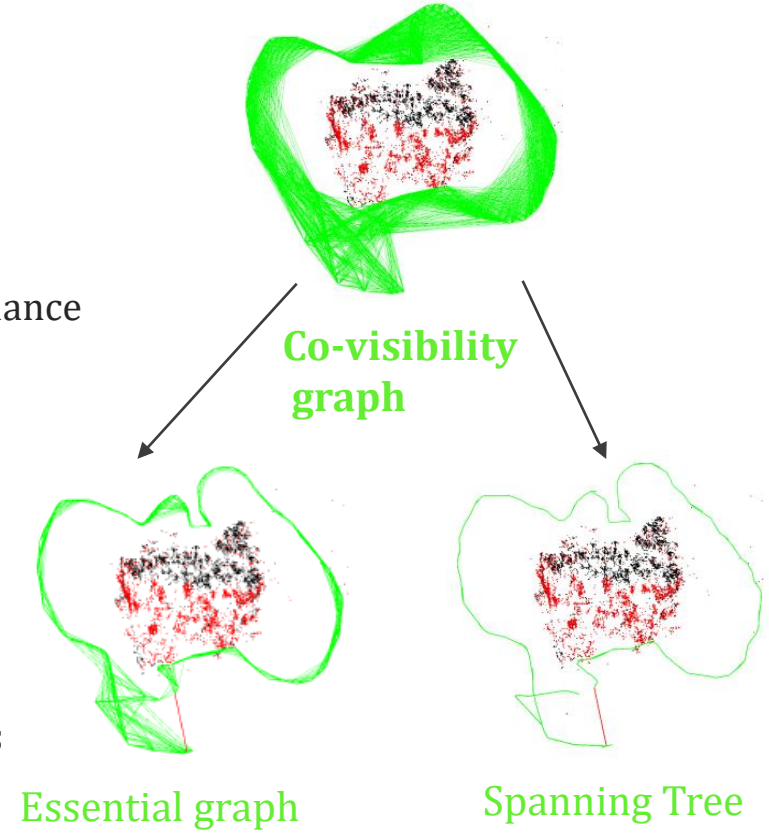
- 3D position X
- Normal vector indicating viewing direction n
- A representative ORB descriptor D
- Max and min observing distances for scale invariance

Key Frames

- camera pose T
- camera intrinsics K
- All ORB features extracted in the frame

Co-visibility graph

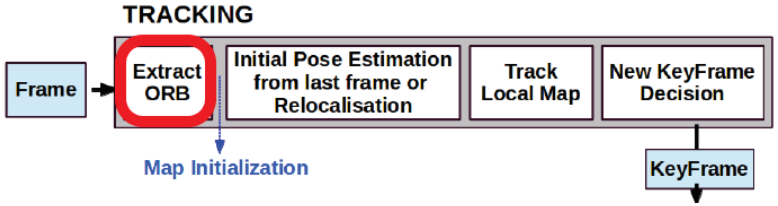
- **Undirected weighted graph**
- Node: keyframe
- Edge: between keyframes that share observations
- Weight θ: the number of common map points



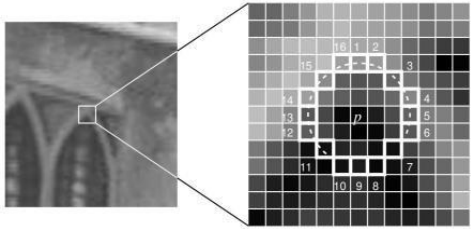
Outline

- Introduction
- Method Description
- Experiments & Results
- Resume

ORB-Extraction



→ oFAST: addition of a fast and accurate orientation component to FAST



FAST-9 keypoint detector

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \qquad \theta = \text{atan2}(m_{01}, m_{10})$$

Patch's moments and the orientation of the patch

→ rBRIEF: constructs a steered version of the keypoints

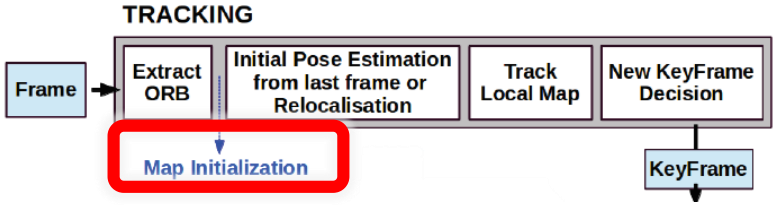
$$\tau(\mathbf{p};\mathbf{x},\mathbf{y}) := \begin{cases} 1 & \text{if } \mathbf{p}(\mathbf{x}) < \mathbf{p}(\mathbf{y}) \\ 0 & \text{otherwise} \end{cases}$$

Binary test on patch intensity
p(x) is the intensity of p at a point x

$$f_{n_d}(\mathbf{p}) := \sum_{1 \le i \le n_d} 2^{i-1}\, \tau(\mathbf{p};\mathbf{x}_i,\mathbf{y}_i) | (\mathbf{x}_i,\mathbf{y}_i) \in \mathbf{S}_\theta$$

Steered BRIEF operator

Map Initialization



Compute score

Low parallax

$$\mathbf{x}_c = \mathbf{H}_{cr}\,\mathbf{x}_r$$

Non-Planar scene

$$\mathbf{x}_c^T\,\mathbf{F}_{cr}\,\mathbf{x}_r = 0$$

Heuristic to select a mode on triangulation of an initial set of map points

$$S_M = \sum_i \left(\rho_M\left(d^2_{cr}(\mathbf{x}_c^i, \mathbf{x}_r^i, M)\right) + \rho_M(d^2_{rc}(\mathbf{x}_c^i, \mathbf{x}_r^i, M))\right)$$

Symmetric transfer error of each model M

$$\rho_M(d^2) = \begin{cases} \Gamma - d^2 & \text{if} \quad d^2 < T_M \\ 0 & \text{if} \quad d^2 \geq T_M \end{cases}$$

the outlier rejection

Model selection

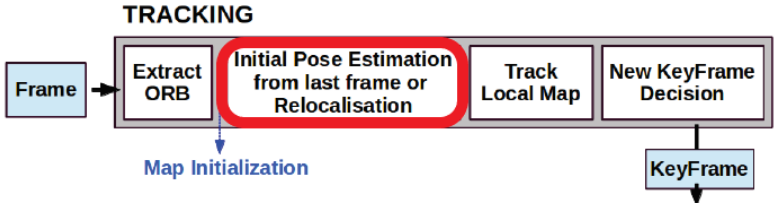
↓

retrieve motion hypotheses

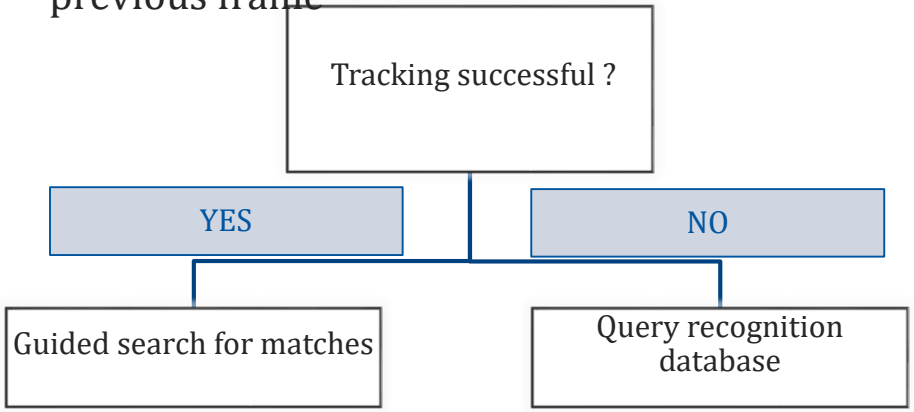
↓

refine with full BA

Tracking Procedure



→ starts with initial feature matching with the previous frame



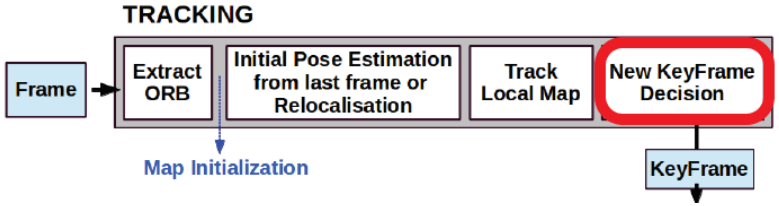
Decision Pose Estimation or Re-Localization

→ optimize the pose using motion-only BA (over set of matches)

$$\{\mathbf{R},\mathbf{t}\} = \underset{\mathbf{R},\mathbf{t}}{\operatorname{argmin}} \sum_{i\in\mathcal{X}} \rho\left(\left\|\mathbf{x}^i_{(\cdot)} - \pi_{(\cdot)}\left(\mathbf{R}\mathbf{X}^i + \mathbf{t}\right)\right\|^2_{\Sigma}\right)$$

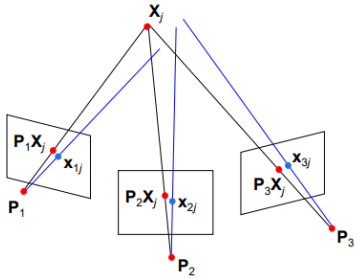
ρ is the robust Huber cost function
Σ the covariance matrix associated to the scale of the keypoint

Tracking Procedure



→ project in frame and search for more map correspondences **locally**

- **discard** map points by mean viewing direction and scale
- compare descriptor and associate with best match



map point reprojection

→ optimize with all map points found

→ Keyframe decision

AS FAST AS POSSIBLE

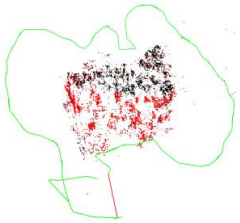
Insertion criteria:

- last global relocalization
- last keyframe insertion (if busy)
- number of tracked points
- visual change criterion

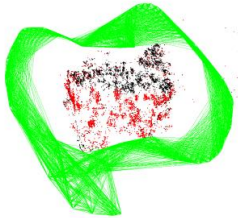
ensures a good relocalization and good tracking

Local Mapping Procedure

→ Updates with **new** keyframe insertion



with most points in common

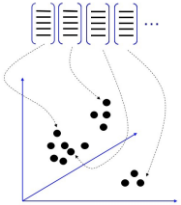


new edges resulting from the shared map points

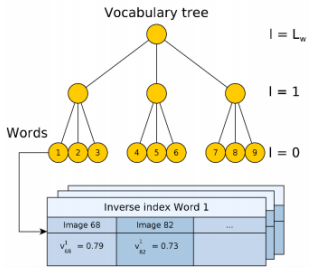
→ Compute the bags of words representation

→ Bag of visual words

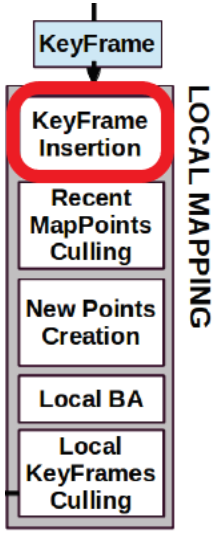
- visual vocabulary is created offline
- binary descriptors traverse the tree
- minimize the Hamming distance



binary clusters over all levels



weight according to its relevance - frequency



Local Mapping Procedure

→ Test keyframe if trackable $\frac{\#\ frames\ where\ match\ is\ found}{\#\ frames\ where\ map\ point\ is\ visible} > 0.25$

→ Triangulating ORB from connected keyframes in the covisablity graph

Parallax, error, scale consistency check

Correspondences are searched, matched, optimized

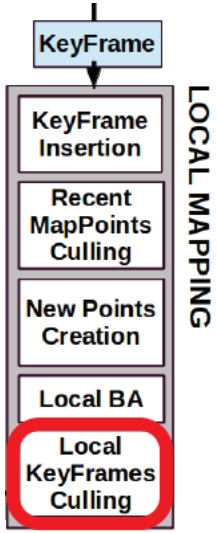
→ Local BA for keyframe and all connected ones

$$\{\mathbf{X}^i, \mathbf{R}_l, \mathbf{t}_l | i \in \mathcal{P}_L, l \in \mathcal{K}_L\} = \underset{\mathbf{X}^i, \mathbf{R}_l, \mathbf{t}_l}{\operatorname{argmin}} \sum_{k \in \mathcal{K}_L \cup \mathcal{K}_F} \sum_{j \in \mathcal{X}_k} \rho\left(E_{kj}\right)$$

optimizes a set of co-visible keyframes and all points seen

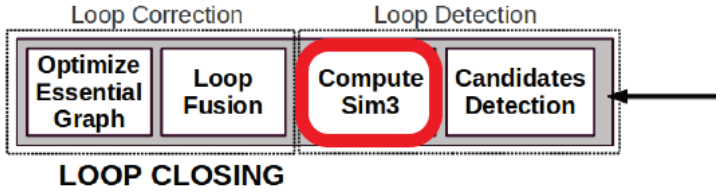
$$E_{kj} = \left\| \mathbf{x}^j_{(\cdot)} - \pi_{(\cdot)}\left(\mathbf{R}_k \mathbf{X}^j + \mathbf{t}_k\right) \right\|^2_{\Sigma}$$

reprojection error of match j in frame k



→ Culling redundant frames

Loop Closing Procedure



→ only for neighbours in co-visability (minimum weight of 30)

→ compute normalizing score

→ check min. 3 consecutive frames

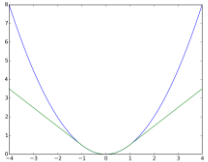
$$s(\mathbf{v}_1, \mathbf{v}_2) = 1 - \frac{1}{2}\left|\frac{\mathbf{v}_1}{|\mathbf{v}_1|} - \frac{\mathbf{v}_2}{|\mathbf{v}_2|}\right|$$

similarity between two bag-of-words vectors v1 and v2

Similarity transform between current keyframe and the loop closure candidate used for geometrical validation of the loop

$$C = \sum_n \left(\rho_h(\mathbf{e}_1^T \mathbf{\Omega}_{1,i}^{-1} \mathbf{e}_1) + \rho_h(\mathbf{e}_2^T \mathbf{\Omega}_{2,j}^{-1} \mathbf{e}_2)\right)$$

Ω are the covariance matrices associated to the scale

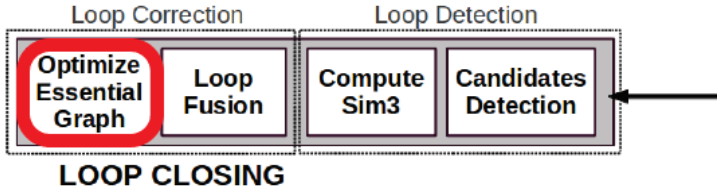


ρ_h Huber loss function (green)

$$\mathbf{e_1} = \mathbf{x}_{1,i} - \pi_1(\mathbf{S}_{12}, \mathbf{X}_{2,j})$$
$$\mathbf{e_2} = \mathbf{x}_{2,j} - \pi_2(\mathbf{S}_{12}^{-1}, \mathbf{X}_{1,i})$$

projection errors of similarities between frame 1 and 2

Loop Closing Procedure



Fuse duplicate points & update edges

Align both sides of the loop, propagation

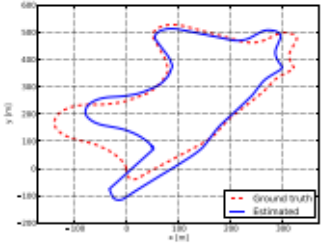
Projection of points into current frame and match points

→ Distribute loop closing error (is a vector in $\mathbb{R}^7$) along the graph

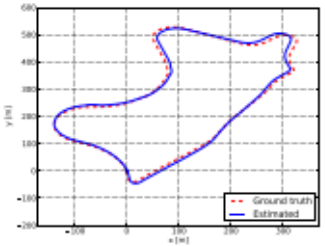
$$\mathbf{e}_{i,j} = \log_{\mathrm{Sim}(3)}(\mathbf{S}_{ij}\,\mathbf{S}_{jw}\,\mathbf{S}_{iw}^{-1}) \qquad C = \sum_{i,j}(\mathbf{e}_{i,j}^T \mathbf{\Lambda}_{i,j}\mathbf{e}_{i,j})$$

Error in an edge of the pose graph

$\Lambda_{i,j}$ is the information matrix of the edge



Without loop closing



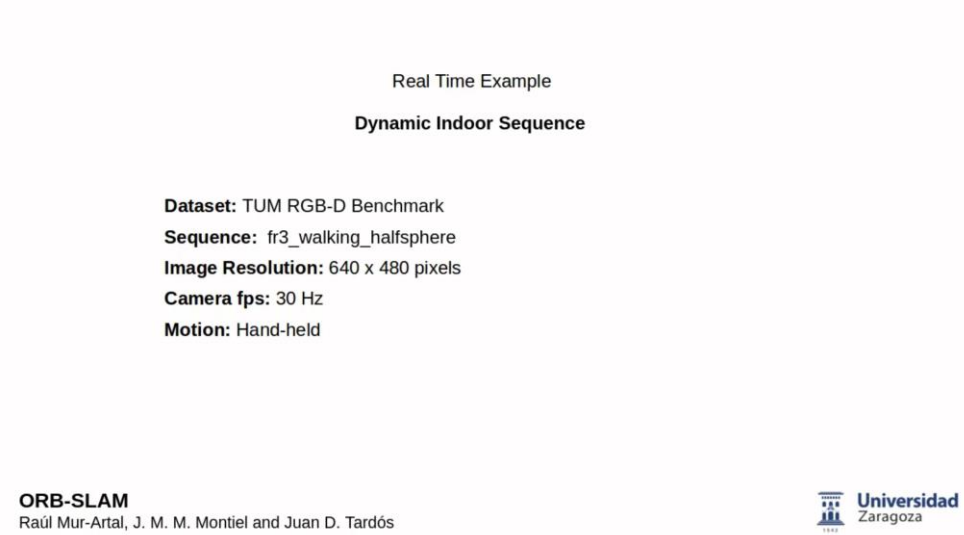
only pose graph optimization

Outline

- Introduction
- Method Description
- Experiments & Results
- Resume

Localization Accuracy

→ TUM RGB-D Benchmark

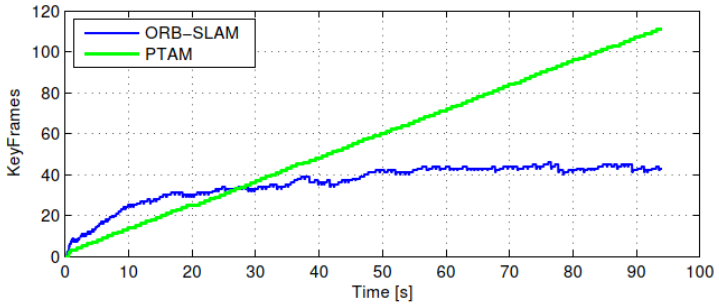


Fr3_walking_halfsphere

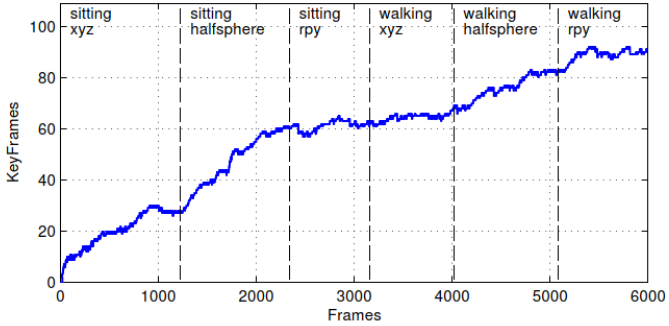
	Absolute KeyFrame Trajectory RMSE (cm)			
	ORB-SLAM	PTAM	LSD-SLAM	RGBD-SLAM
fr1_xyz	0.90	1.15	9.00	1.34 (1.34)
fr2_xyz	0.30	**0.20**	2.15	2.61 (1.42)
fr1_floor	2.99	X	38.07	3.51 (3.51)
fr1_desk	1.69	X	10.65	2.58 (2.52)
fr2_360_kidnap	3.81	**2.63**	X	393.3 (100.5)
fr2_desk	0.88	X	4.57	9.50 (3.94)
fr3_long_office	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	-

Results for ORB-SLAM, PTAM and LSD-SLAM are the median over 5 executions in each sequence

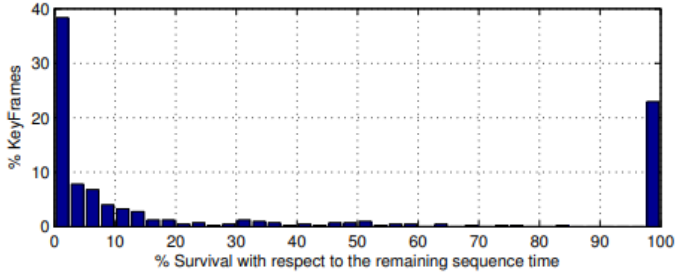
Map Expansion / Lifetime



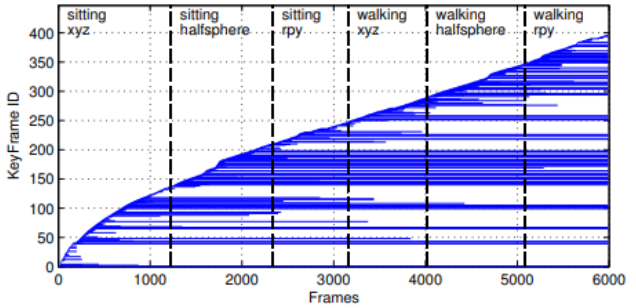
Map expansion in static scence (same content, different viewpoint)



Evolution of the number of keyframes in the map



Histogram of the survival time of all spawned keyframes with respect to the remaining time of the experiment



survival time of all spawned keyframes in similar environment, different viewpoints

Loop Closing

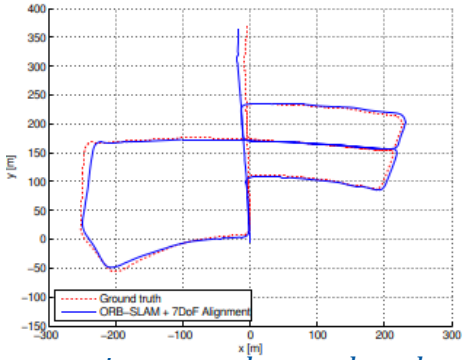
Real Time Example

Large Scale Outdoors with multiple Loops

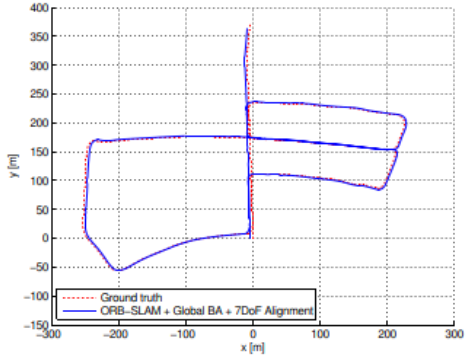
Dataset: KITTI (Odometry benchmark)
Sequence: 05
Dimensions: 479 x 426 meters
Image Resolution: 1241 x 376 pixels
Camera fps: 10 Hz
Motion: Car

ORB-SLAM
Raúl Mur-Artal, J. M. M. Montiel and Juan D. Tardós

Universidad Zaragoza



trajectory and ground truth



trajectory after 20 iterations of full BA

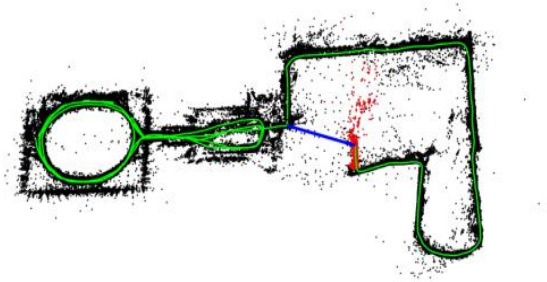
Performance

→ New College Dataset

With several loops and fast rotations

Thread	Operation	Median (ms)	Mean (ms)	Std (ms)
TRACKING	ORB extraction	11.10	11.42	1.61
	Initial Pose Est.	3.38	3.45	0.99
	Track Local Map	14.84	16.01	9.98
	Total	30.57	31.60	10.39
LOCAL MAPPING	KeyFrame Insertion	10.29	11.88	5.03
	Map Point Culling	0.10	3.18	6.70
	Map Point Creation	66.79	72.96	31.48
	Local BA	296.08	360.41	171.11
	KeyFrame Culling	8.07	15.79	18.98
	Total	383.59	464.27	217.89

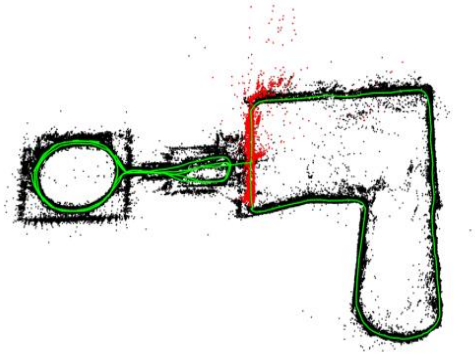
Tracking and Mapping Times



Map before loop closure

			Loop Detection (ms)		Loop Correction (s)		
Loop	KeyFrames	Essential Graph Edges	Candidates Detection	Similarity Transformation	Fusion	Essential Graph Optimization	Total (s)
1	287	1347	4.71	20.77	0.20	0.26	0.51
2	1082	5950	4.14	17.98	0.39	1.06	1.52
3	1279	7128	9.82	31.29	0.95	1.26	2.27
4	2648	12547	12.37	30.36	0.97	2.30	3.33
5	3150	16033	14.71	41.28	1.73	2.80	4.60
6	4496	21797	13.52	48.68	0.97	3.62	4.69

Loop Closing Times in New College Dataset



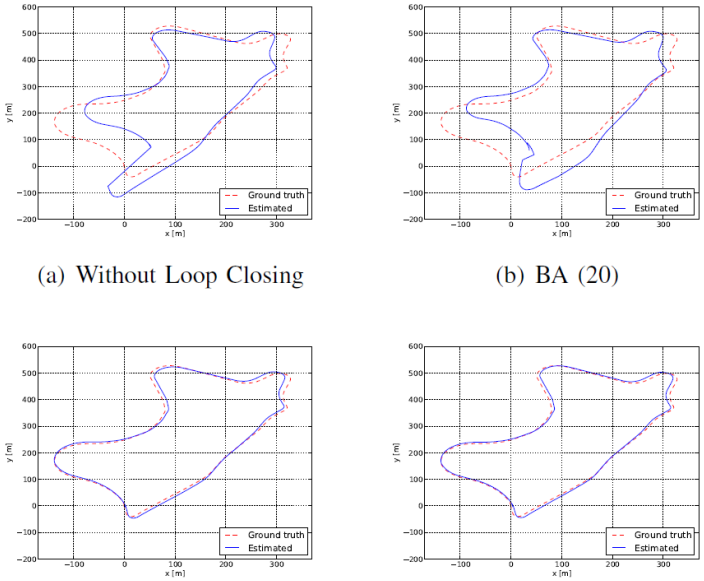
Map after a loop closure

Experiments – Loop Closing Strategies

KITTI dataset

Method	Time (s)	Pose Graph Edges	RMSE (m)
-	-	-	48.77
BA (20)	14.64	-	49.90
BA (100)	72.16	-	18.82
EG (200)	0.38	890	8.84
EG (100)	0.48	1979	8.36
EG (50)	0.59	3583	8.95
EG (15)	0.94	6663	8.88
EG (100) + BA (20)	13.40	1979	7.22

LOOP CLOSING STRATEGIES IN KITTI 09



(a) Without Loop Closing (b) BA (20)

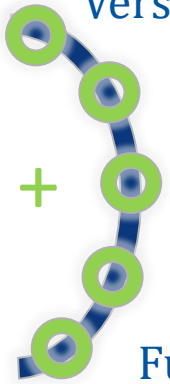
(c) EG (100) (d) EG (100) + BA (20)

Comparison of different loop closing strategies in KITTI 09

Outline

- Introduction
- Method Description
- Experiments & Results
- Resume

Resume



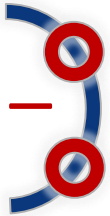
Versatility: indoor/outdoor scene, car, robot, hand-held motions

Flexible map expansion

\+ Essential graph optimization shows fast convergence

Real time accurate tracking and mapping without GPU

Full system/application

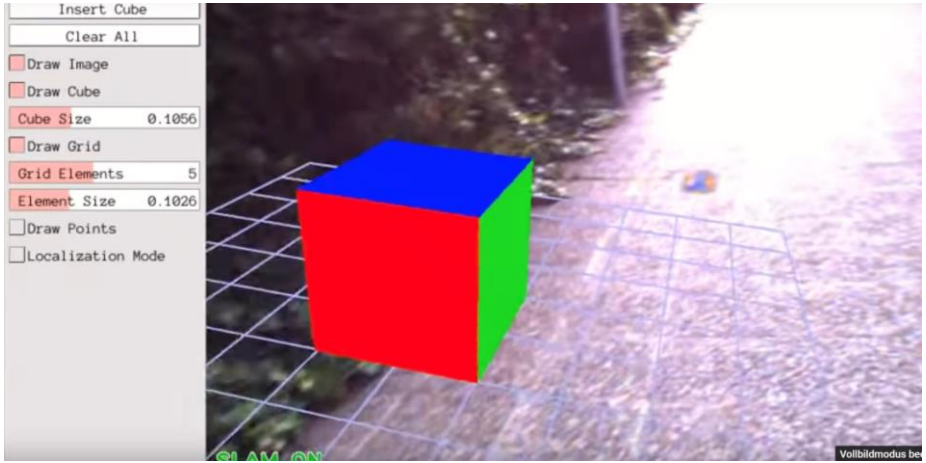


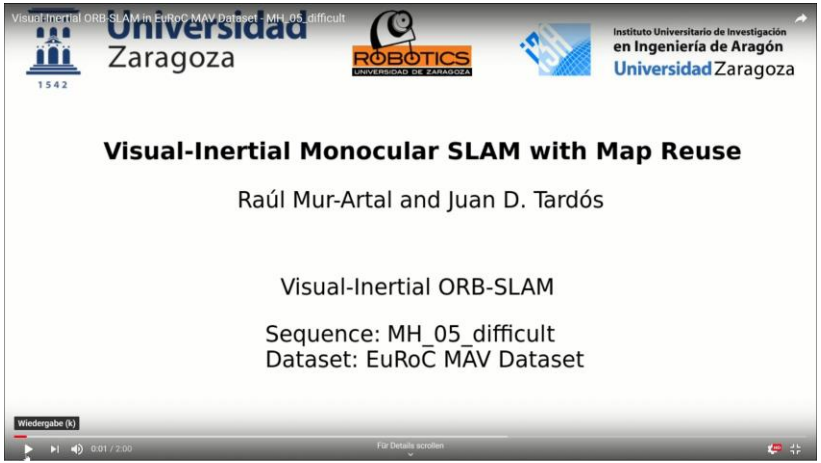
\– Limited application domains according to reconstruction accuracy

Algorithm may not initialize

Related Works

- Extension ORB-SLAM 2 and open-source SLAM framework
- Semi-Dense Mapping from Highly Accurate Feature-Based Monocular SLAM
- Visual-Inertia Monocular SLAM with Map Reuse

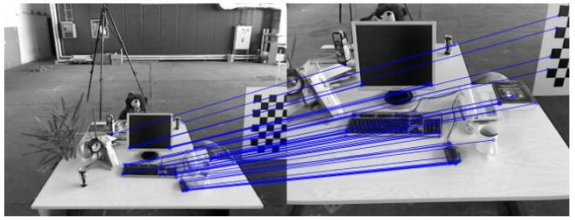


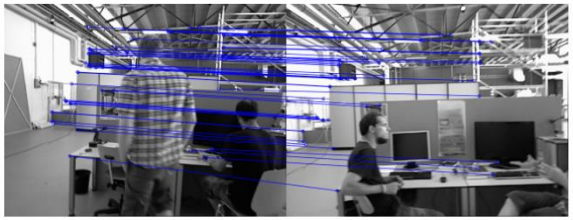


Back up Slides

Relocalization

→ TUM RGB-D Benchmark

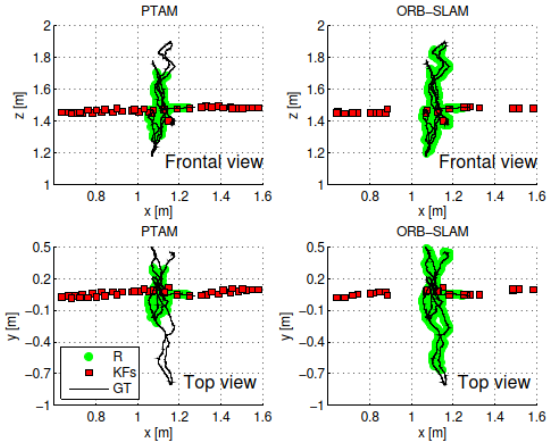




ORB-SLAM finds enough inlier correspondences supporting the similarity transformation for the relocalization

System	Initial Map		Relocalization		
	KFs	RMSE (cm)	Recall (%)	RMSE (cm)	Max. Error (cm)
fr2_xyz. 2769 frames to relocalize					
PTAM	37	0.19	34.9	0.26	1.52
ORB-SLAM	24	0.19	**78.4**	0.38	1.67
fr3_walking_xyz. 859 frames to relocalize					
PTAM	34	0.83	0.0	-	-
ORB-SLAM	31	0.82	**77.9**	1.32	4.95

Relocalization in challenging environment in comparison to PTAM

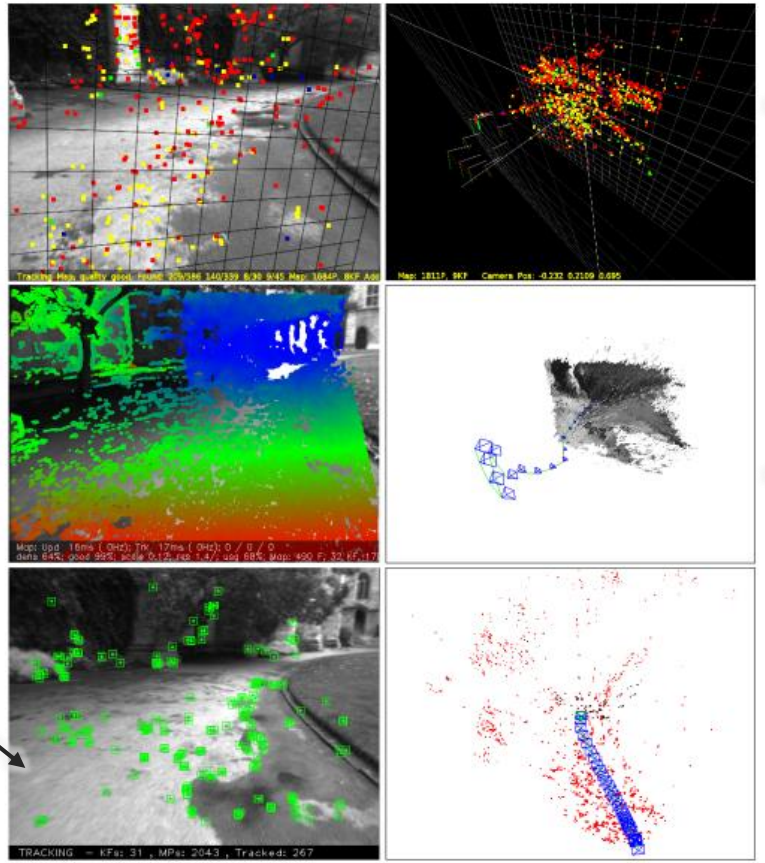


Successful relocalizations in green

Initialization

ORB-SLAM

automatically initialized from the fundamental matrix when it has detected enough parallax



PTAM and LSD-SLAM

Initialize a corrupted planar solution

NewCollege robot sequence