Should we still do sparse-feature based SLAM?

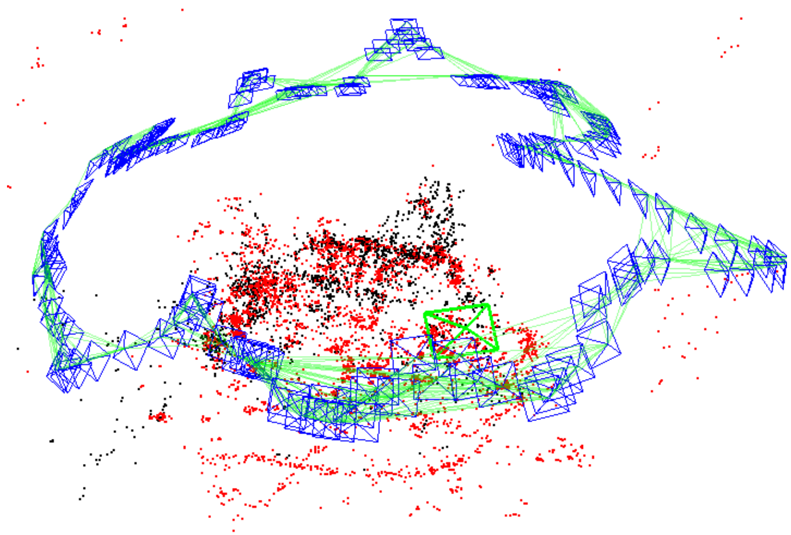
Raúl Mur Artal
PhD. student, University of Zaragoza

ROBOTICS UNIVERSIDAD DE ZARAGOZA

Universidad Zaragoza 1542

Feature-based/Direct SLAM

Feature-Based SLAM

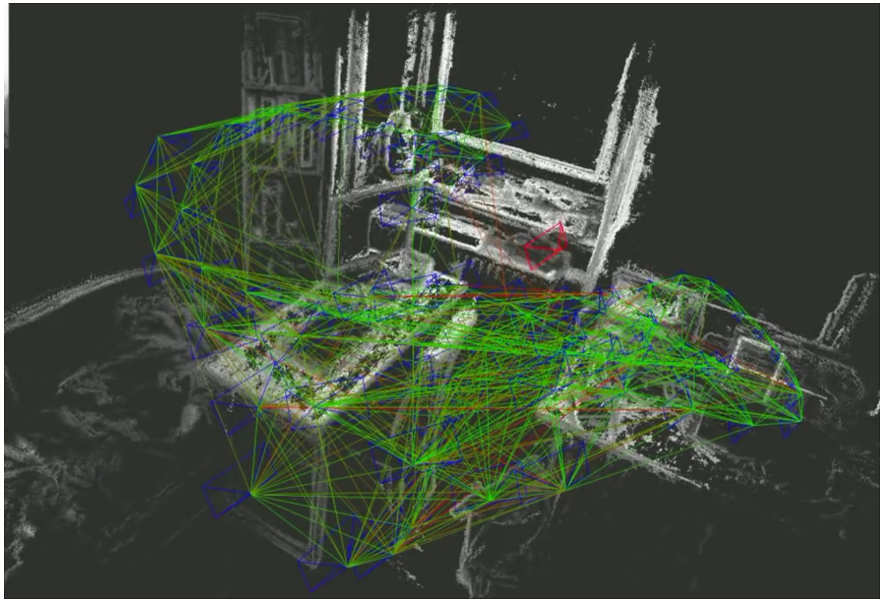


Minimize **Feature Reprojection Error**

Sparse Reconstruction

PTAM, ORB-SLAM

Direct SLAM



Minimize **Photometric Error**

Semi Dense / Dense Reconstruction

DTAM, LSD-SLAM, DPPTAM

Outline

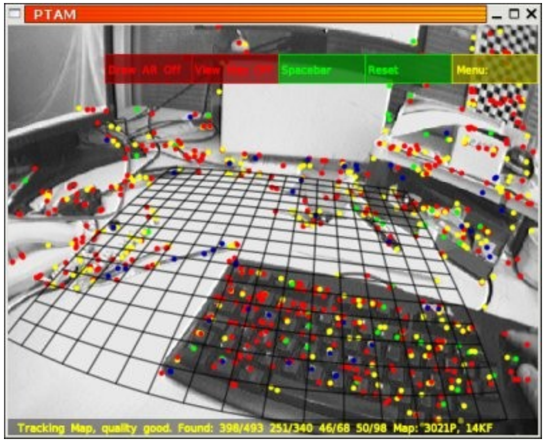
1) **Monocular ORB-SLAM**
(Mur-Artal, Montiel, Tardos T-RO 2015)

2) **Monocular Semi Dense Mapping**
(Mur-Artal, Tardos, RSS 2015)

3) **Stereo/RGB-D ORB-SLAM**
(unpublished)

Motivation

Parallel Tracking and Mapping (PTAM)
G. Klein and D. Murray
Parallel Tracking and Mapping for Small AR Workspaces
ISMAR 2007.

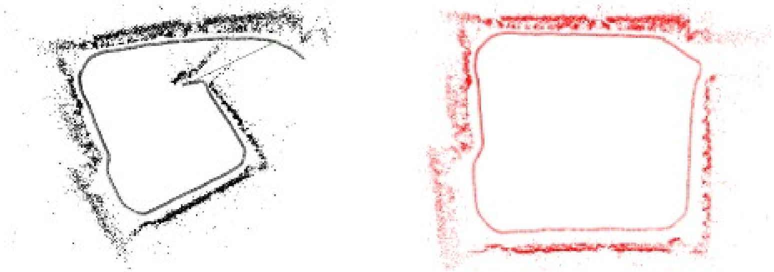


- ✓ **First keyframe-based monocular SLAM**
- ✓ **Bundle adjustment is possible in real-time**

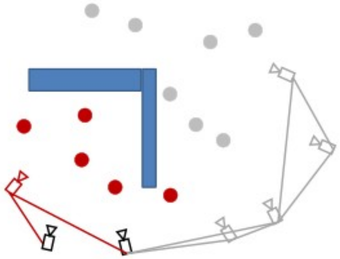
- ✗ **Relocalisation with little invariance to viewpoint**
- ✗ **No loop closing/detection mechanism**
- ✗ **Small scale operation**
- ✗ **User intervention for map initialization and a dominant plane**

Motivation

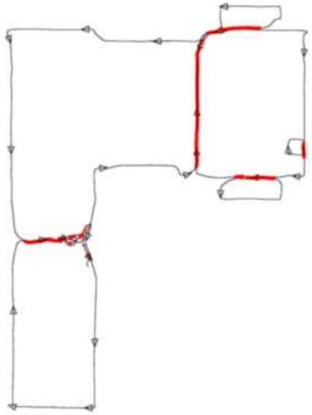
Scale Drift-Aware Loop Closing
H. Strasdat, J. M. M. Montiel and A. J. Davison
Scale Drift-Aware Large Scale Monocular SLAM
RSS 2010



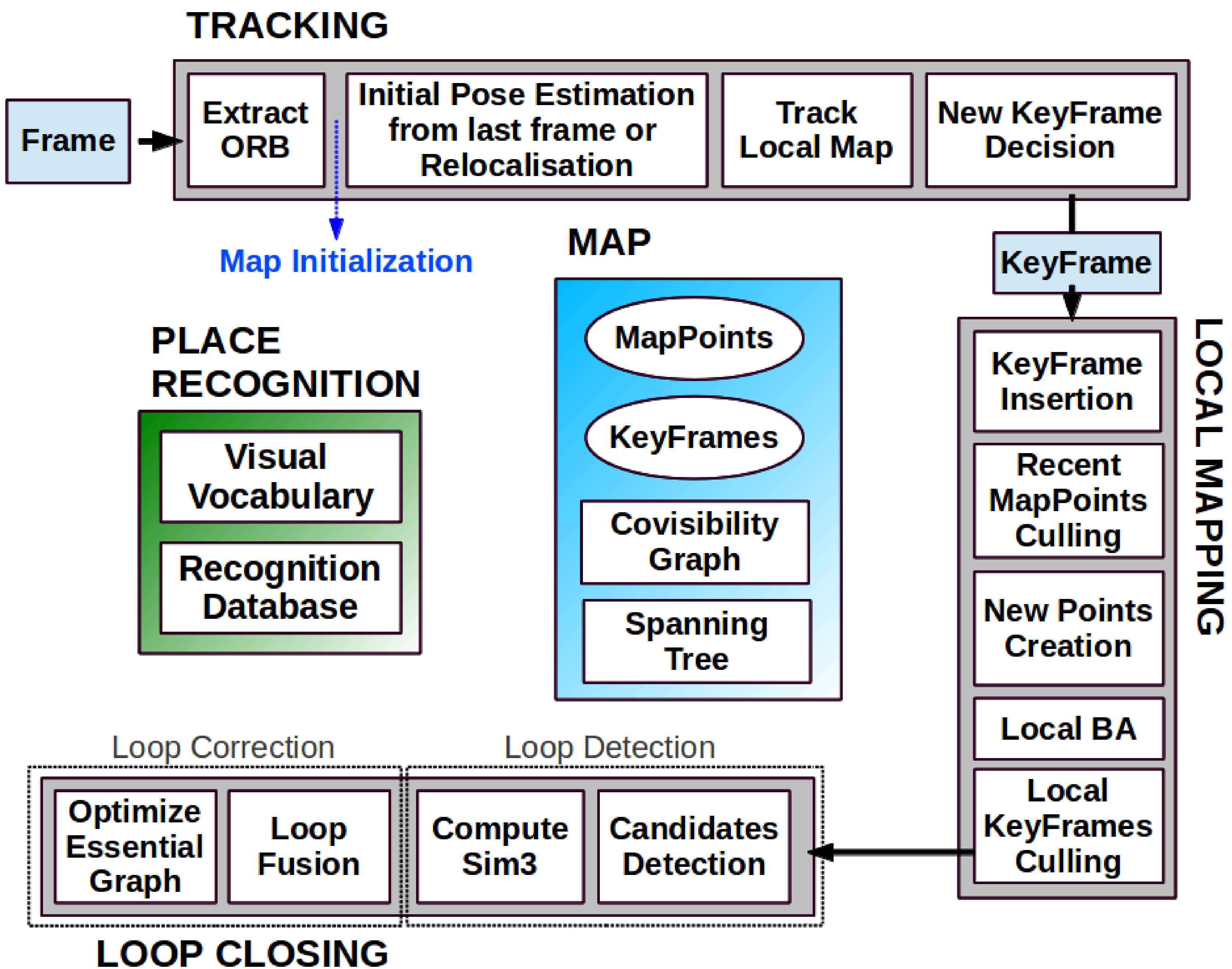
Covisibility Graph
H. Strasdat, A. J. Davison, J. M. M. Montiel , K. Konolige
Double Window Optimization for Constant Time Visual SLAM
ICCV 2011



Bags of Binary Words (DBoW)
D. Gálvez-López and J. D. Tardós
Bags of Binary Words for Fast Place Recognition in Image Sequences
T-RO 2012



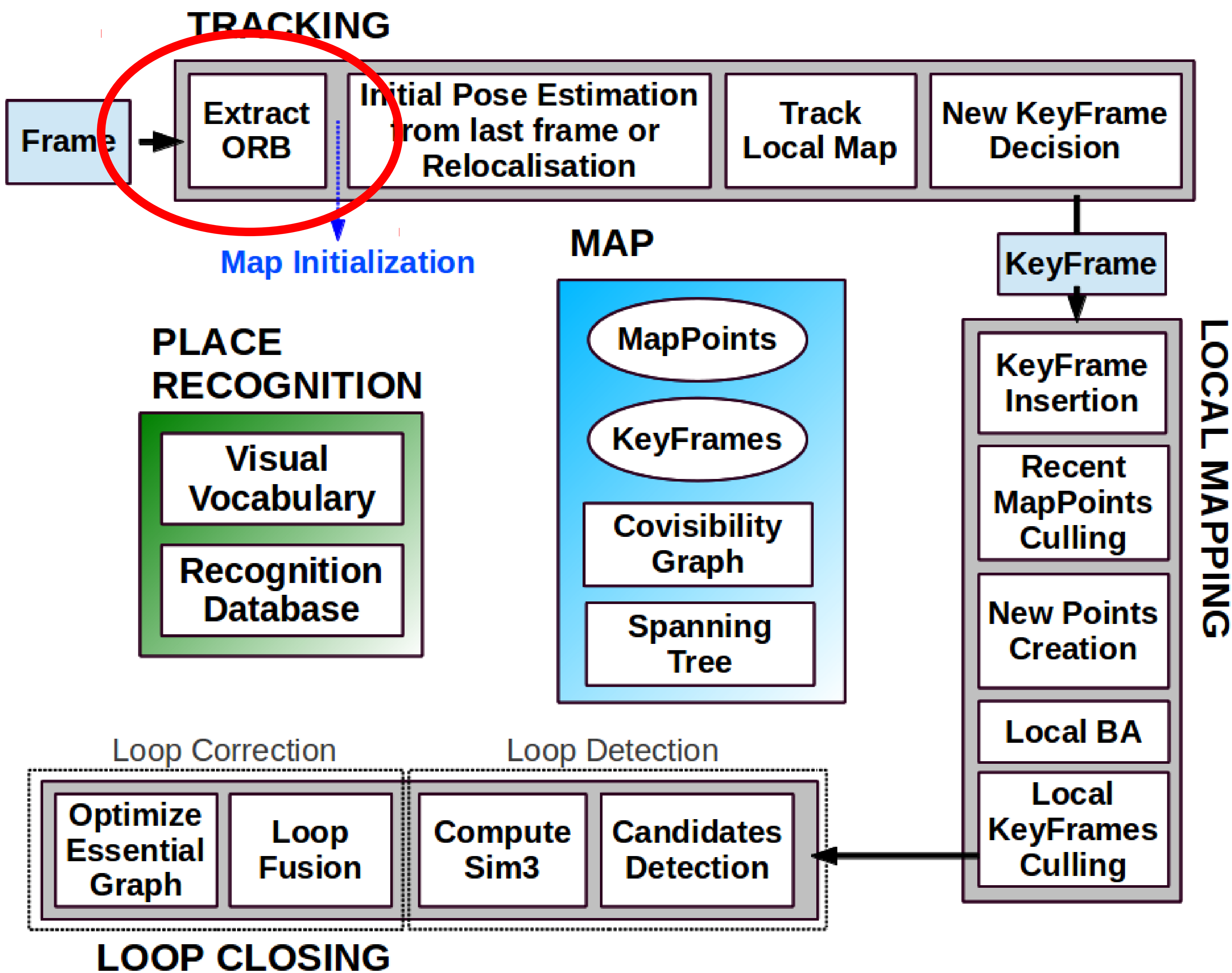
ORB-SLAM



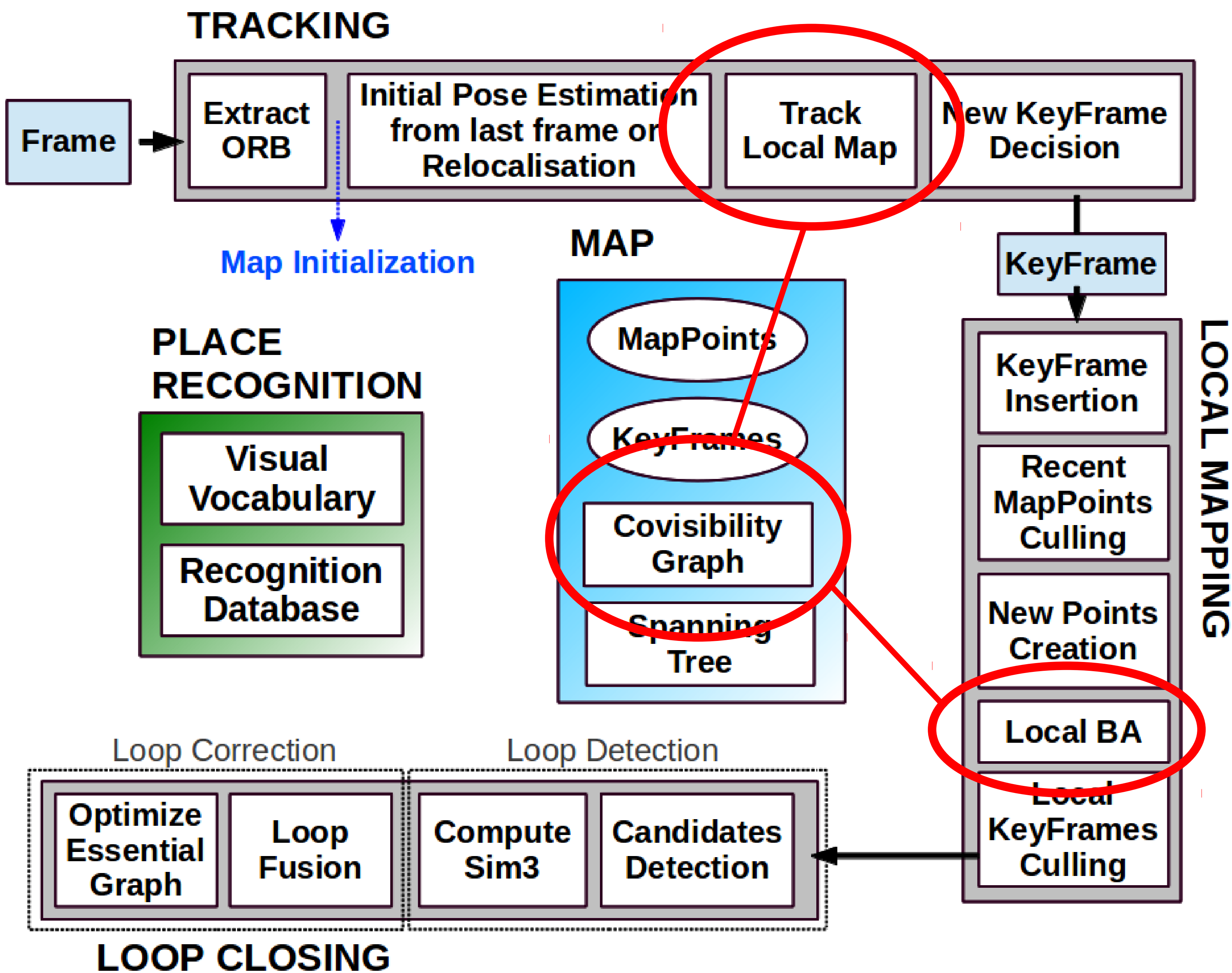
TRACKING
Frame
Extract ORB
Initial Pose Estimation from last frame or Relocalisation
Track Local Map
New KeyFrame Decision
Map Initialization
MAP
KeyFrame
MapPoints
KeyFrames
Covisibility Graph
Spanning Tree
PLACE RECOGNITION
Visual Vocabulary
Recognition Database
LOCAL MAPPING
KeyFrame Insertion
Recent MapPoints Culling
New Points Creation
Local BA
Local KeyFrames Culling
Loop Correction
Loop Detection
Optimize Essential Graph
Loop Fusion
Compute Sim3
Candidates Detection
LOOP CLOSING

ORB-SLAM

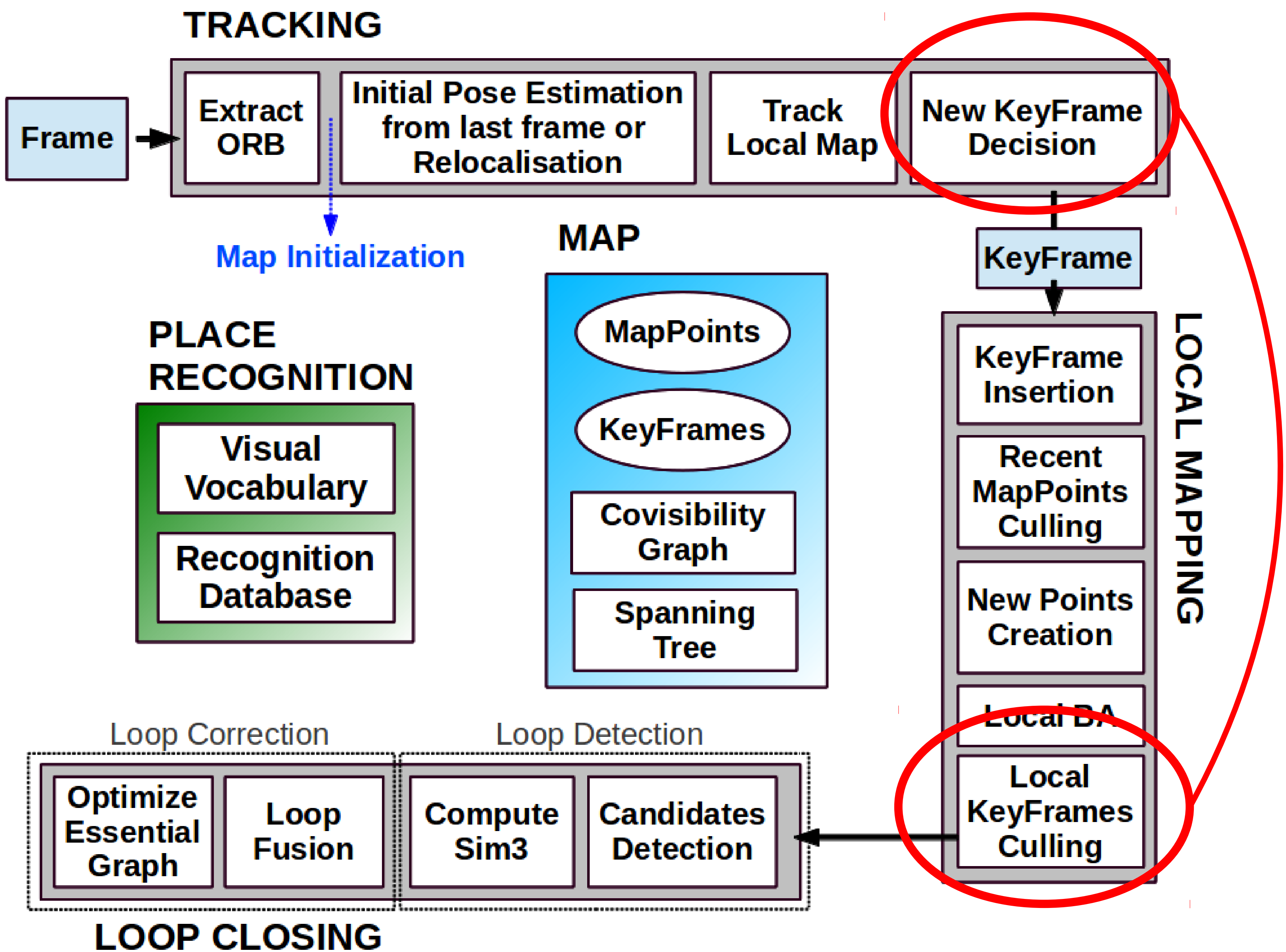
Universidad Zaragoza
1542



ORB-SLAM

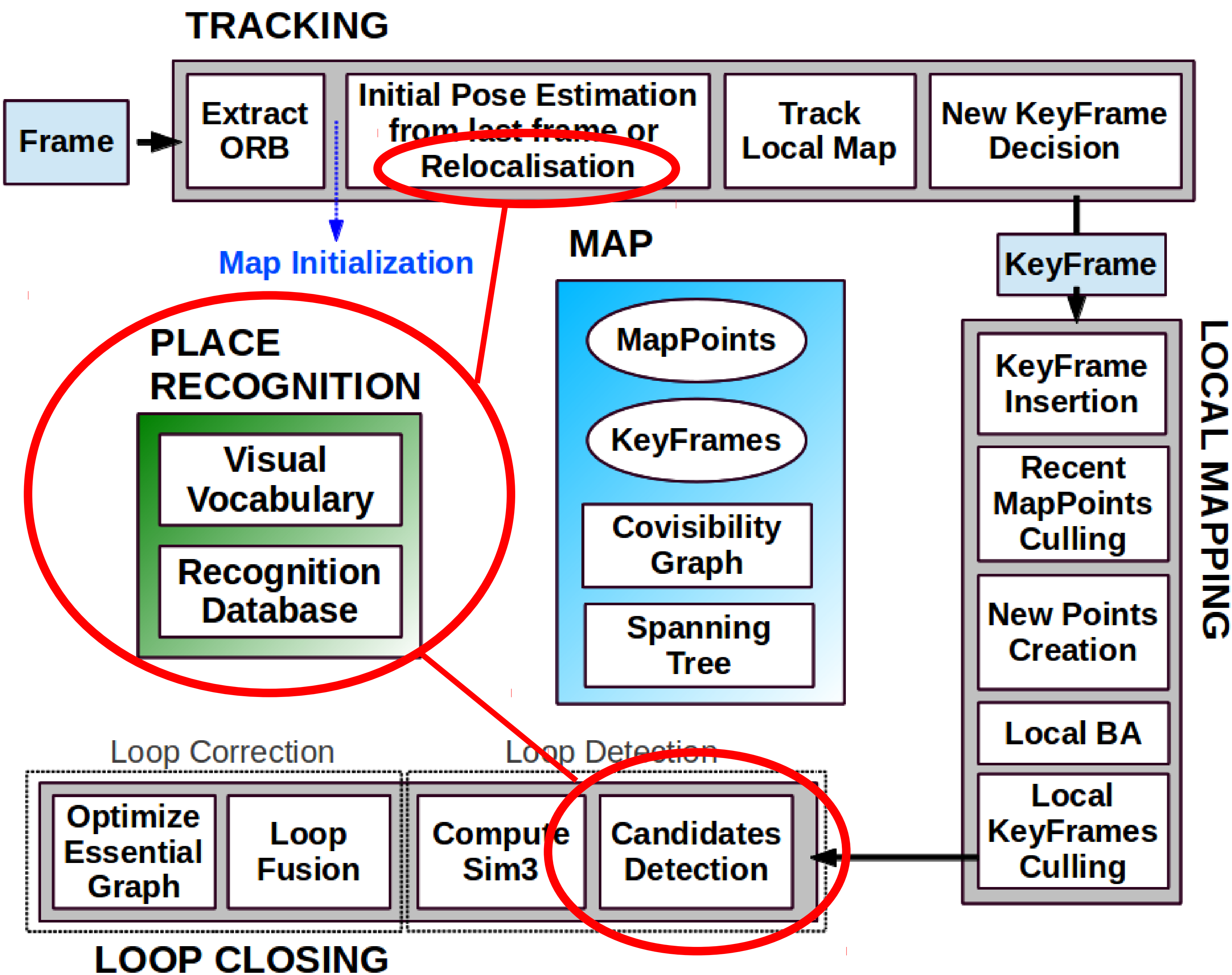

TRACKING
Frame
Extract ORB
Initial Pose Estimation from last frame or Relocalisation
Track Local Map
New KeyFrame Decision
Map Initialization
MAP
KeyFrame
MapPoints
KeyFrames
Covisibility Graph
Spanning Tree
PLACE RECOGNITION
Visual Vocabulary
Recognition Database
LOCAL MAPPING
KeyFrame Insertion
Recent MapPoints Culling
New Points Creation
Local BA
Local KeyFrames Culling
Loop Correction
Loop Detection
Optimize Essential Graph
Loop Fusion
Compute Sim3
Candidates Detection
LOOP CLOSING

ORB-SLAM



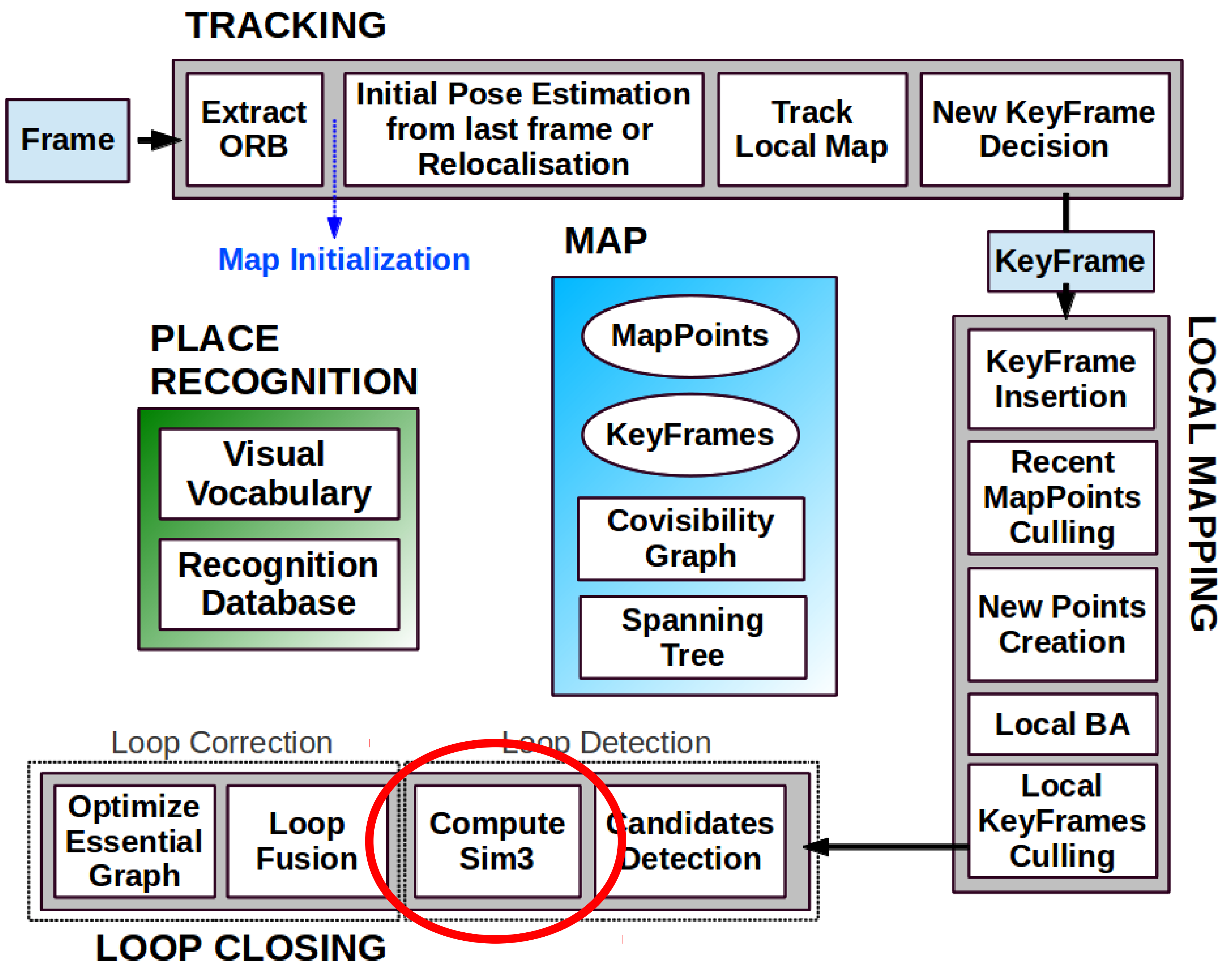
TRACKING
Frame
Extract ORB
Initial Pose Estimation from last frame or Relocalisation
Track Local Map
New KeyFrame Decision
Map Initialization
MAP
KeyFrame
MapPoints
KeyFrames
Covisibility Graph
Spanning Tree
PLACE RECOGNITION
Visual Vocabulary
Recognition Database
LOCAL MAPPING
KeyFrame Insertion
Recent MapPoints Culling
New Points Creation
Local BA
Local KeyFrames Culling
Loop Correction
Loop Detection
Optimize Essential Graph
Loop Fusion
Compute Sim3
Candidates Detection
LOOP CLOSING

ORB-SLAM

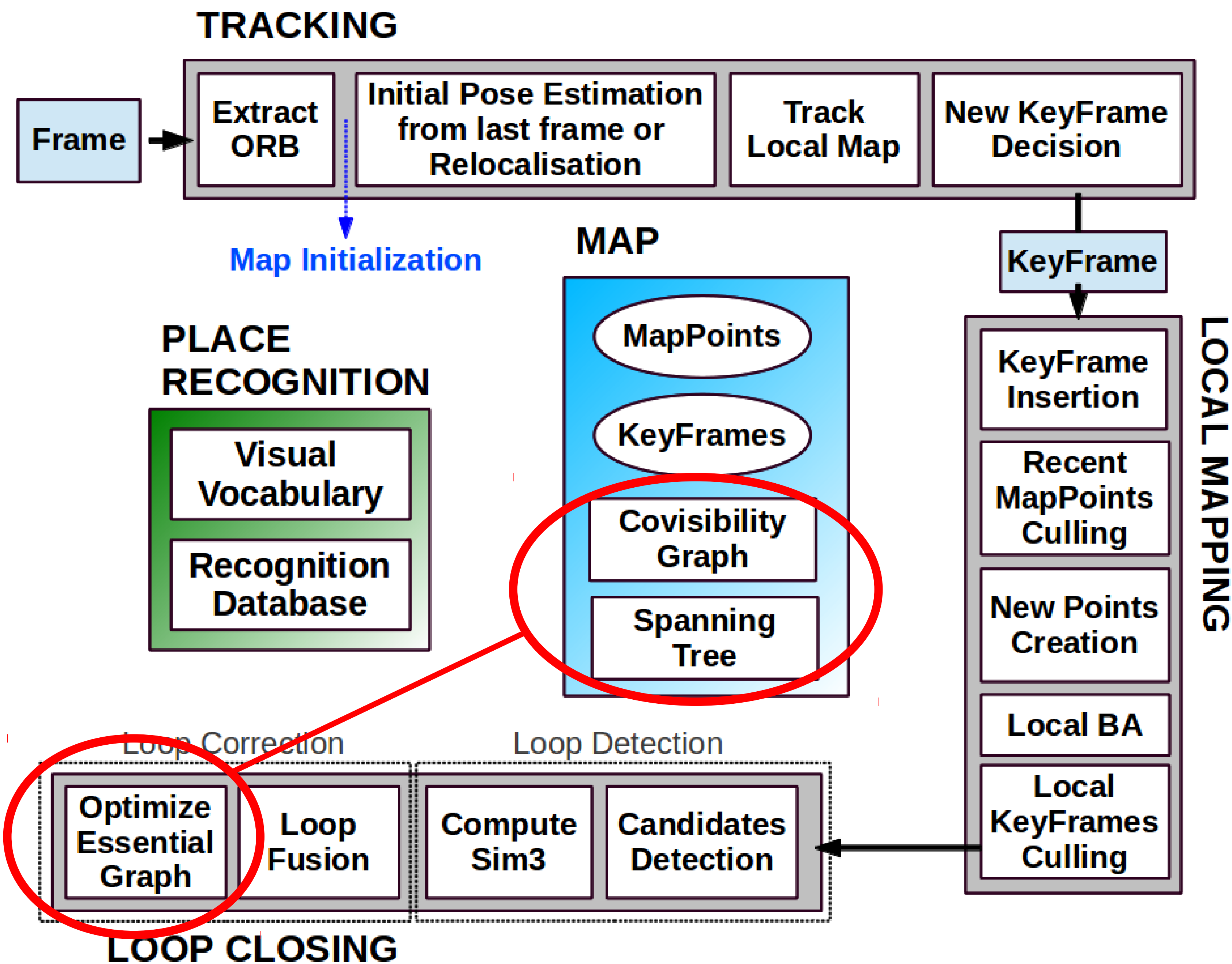


TRACKING
Frame
Extract ORB
Initial Pose Estimation from last frame or Relocalisation
Track Local Map
New KeyFrame Decision
Map Initialization
MAP
KeyFrame
MapPoints
KeyFrames
Covisibility Graph
Spanning Tree
PLACE RECOGNITION
Visual Vocabulary
Recognition Database
LOCAL MAPPING
KeyFrame Insertion
Recent MapPoints Culling
New Points Creation
Local BA
Local KeyFrames Culling
Loop Correction
Loop Detection
Optimize Essential Graph
Loop Fusion
Compute Sim3
Candidates Detection
LOOP CLOSING

ORB-SLAM


TRACKING
Frame
Extract ORB
Initial Pose Estimation from last frame or Relocalisation
Track Local Map
New KeyFrame Decision
Map Initialization
MAP
KeyFrame
PLACE RECOGNITION
Visual Vocabulary
Recognition Database
MapPoints
KeyFrames
Covisibility Graph
Spanning Tree
LOCAL MAPPING
KeyFrame Insertion
Recent MapPoints Culling
New Points Creation
Local BA
Local KeyFrames Culling
Loop Correction
Loop Detection
Optimize Essential Graph
Loop Fusion
Compute Sim3
Candidates Detection
LOOP CLOSING

ORB-SLAM


TRACKING
Frame
Extract ORB
Initial Pose Estimation from last frame or Relocalisation
Track Local Map
New KeyFrame Decision
Map Initialization
MAP
KeyFrame
MapPoints
KeyFrames
Covisibility Graph
Spanning Tree
PLACE RECOGNITION
Visual Vocabulary
Recognition Database
LOCAL MAPPING
KeyFrame Insertion
Recent MapPoints Culling
New Points Creation
Local BA
Local KeyFrames Culling
Loop Correction
Loop Detection
Optimize Essential Graph
Loop Fusion
Compute Sim3
Candidates Detection
LOOP CLOSING

Automatic Map Initialization

ORB-SLAM

Model Selection

Homography
(Planar, Low Parallax)

Fundamental Matrix
(General)

https://youtu.be/UH8KXK8U_Ko

PTAM

Tracking I: Occlusion Handling

Covisibility Graph

Track only a local map (potentially visible)

MapPoint Mean Viewing Direction

Do not project if further than 60º

ORB-SLAM

https://youtu.be/UH8KXK8U_Ko?t=16s

PTAM

Tracking II: KeyFrame Insertion

ORB-SLAM

Survival of the Fittest
KeyFrame Selection

Little Restrictive Insertion
(no distance threshold)

Restrictive Culling
(redundant keyframes
detection)

PTAM

https://youtu.be/UH8KXK8U_Ko?t=46s

Relocalisation

ORB-SLAM

PTAM

<u>Bags of Binary Words</u>

Same ORB than
Tracking and Mapping

High Viewpoint
Invariance
(ORB)

https://youtu.be/UH8KXK8U_Ko?t=1m11s

ORB-SLAM

KITTI Dataset. Sequence 00

https://youtu.be/8DISRmsO2YQ

Universidad Zaragoza

TUM RGB-D Benchmark

RMS KeyFrame Position Error (cm)

Median over 5 executions

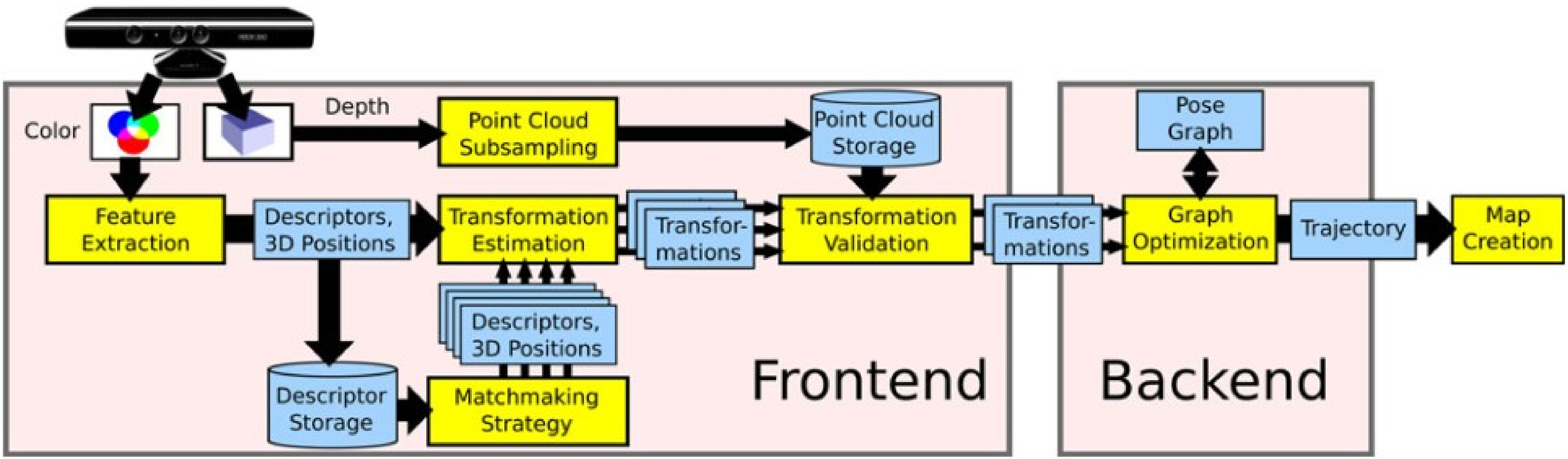
	ORB-SLAM	PTAM
fr1_xyz	**0.90**	1.15
fr2_xyz	0.30	**0.20**
fr1_floor	**2.99**	
fr1_desk	**1.69**	
fr2_360_kidnap	3.81	**2.63**
fr2_desk	**0.88**	
fr3_long_office	**3.45**	
fr3_nstr_tex_near	**1.39**	2.74
fr3_str_tex_far	**0.77**	0.93
fr3_str_tex_near	1.58	**1.04**
fr2_desk_person	**0.63**	
fr3_sit_xyz	**0.79**	0.83
fr3_sit_halfsph	**1.34**	
fr3_walk_xyz	**1.24**	
fr3_walk_halfsph	**1.74**	

Tracking failure

ROBOTICS UNIVERSIDAD DE ZARAGOZA

Comparison with RGBD-SLAM

Use depth information!



F. Endres, J. Hess, J. Sturm, D. Cremers and W. Burgard
3-D Mapping with an RGB-D Camera
IEEE Transaction on Robotics, 2014.

Universidad Zaragoza

TUM RGB-D Benchmark

RMSE (cm)

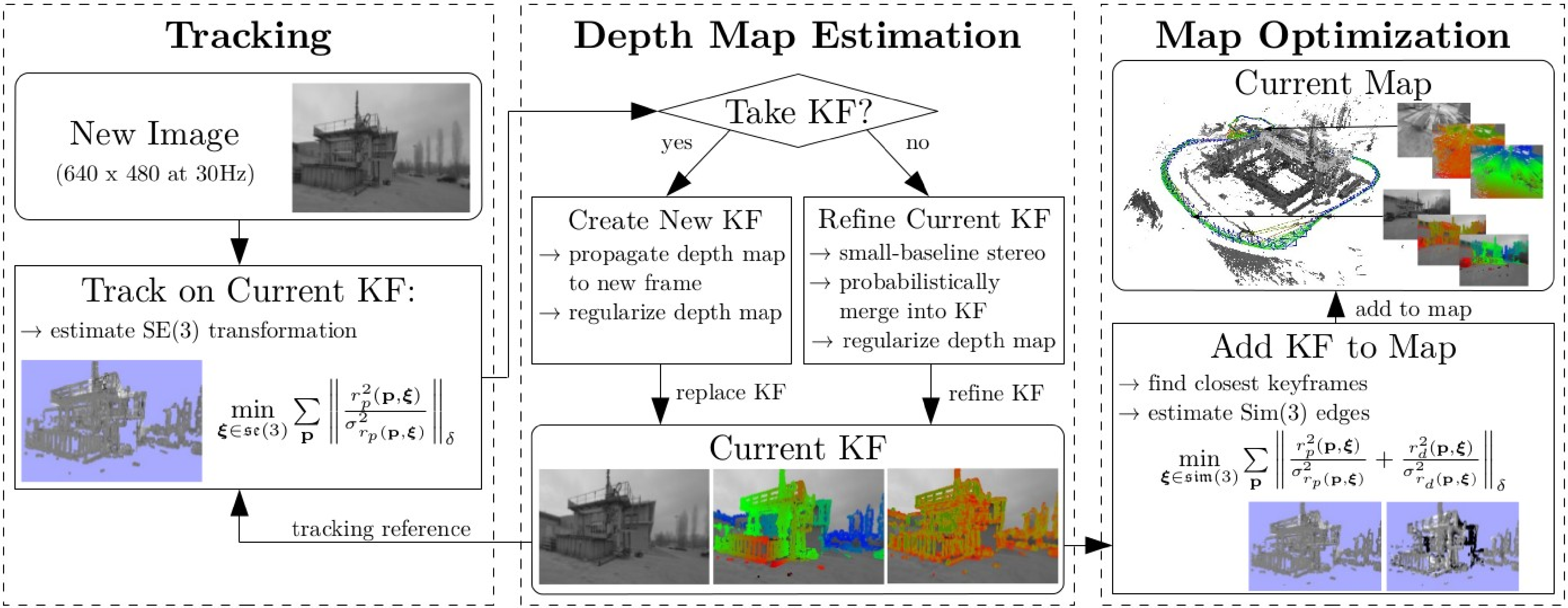
	ORB-SLAM	PTAM	RGBD-SLAM
fr1_xyz	**0.90**	1.15	1.34
fr2_xyz	0.30	**0.20**	1.42
fr1_floor	**2.99**		3.51
fr1_desk	**1.69**		2.52
fr2_360 _kidnap	3.81	**2.63**	100.5
fr2_desk	**0.88**		3.94
fr3_long _office	**3.45**		-
fr3_nstr_ tex_near	**1.39**	2.74	-
fr3_str_ tex_far	**0.77**	0.93	-
fr3_str_ tex_near	1.58	**1.04**	-
fr2_desk _person	**0.63**		2.00
fr3_sit_ xyz	**0.79**	0.83	-
fr3_sit_ _halfsph	**1.34**		-
fr3_walk _xyz	**1.24**		-
fr3_walk _halfsph	**1.74**		-

RGB-D SLAM trajectories from the benchmark website

ROBOTICS UNIVERSIDAD DE ZARAGOZA

Comparison with LSD-SLAM

Use directly pixel intensities!



J. Engel, T. Schöps, D. Cremers
LSD-SLAM: Large-Scale Direct Monocular SLAM
European Conference on Computer Vision (ECCV), 2014.

State-of-the-art in Direct SLAM

TUM RGB-D Benchmark

RMSE (cm)

	ORB-SLAM	PTAM	RGBD-SLAM	LSD-SLAM
fr1_xyz	**0.90**	1.15	1.34	9.00
fr2_xyz	0.30	**0.20**	1.42	2.15
fr1_floor	**2.99**		3.51	38.07
fr1_desk	**1.69**		2.52	10.65
fr2_360_kidnap	3.81	**2.63**	100.5	
fr2_desk	**0.88**		3.94	4.57
fr3_long_office	**3.45**		-	38.53
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**	-	
fr2_desk_person	**0.63**		2.00	31.73
fr3_sit_xyz	**0.79**	0.83	-	7.73
fr3_sit_halfsph	**1.34**		-	5.87
fr3_walk_xyz	**1.24**		-	12.44
fr3_walk_halfsph	**1.74**		-	

Comparison with LSD-SLAM

ORB-SLAM

fr3/structure_
texture_near

https://youtu.be/UH8KXK8U_Ko?t=1m40s

LSD-SLAM

Why should we still use features?

Robustness

- **Reliable two-view monocular initialization**
- **Good invariance to viewpoint and illumination**
- **Less affected by auto-gain and auto-exposure**
- **Less affected by dynamic elements**

Accuracy

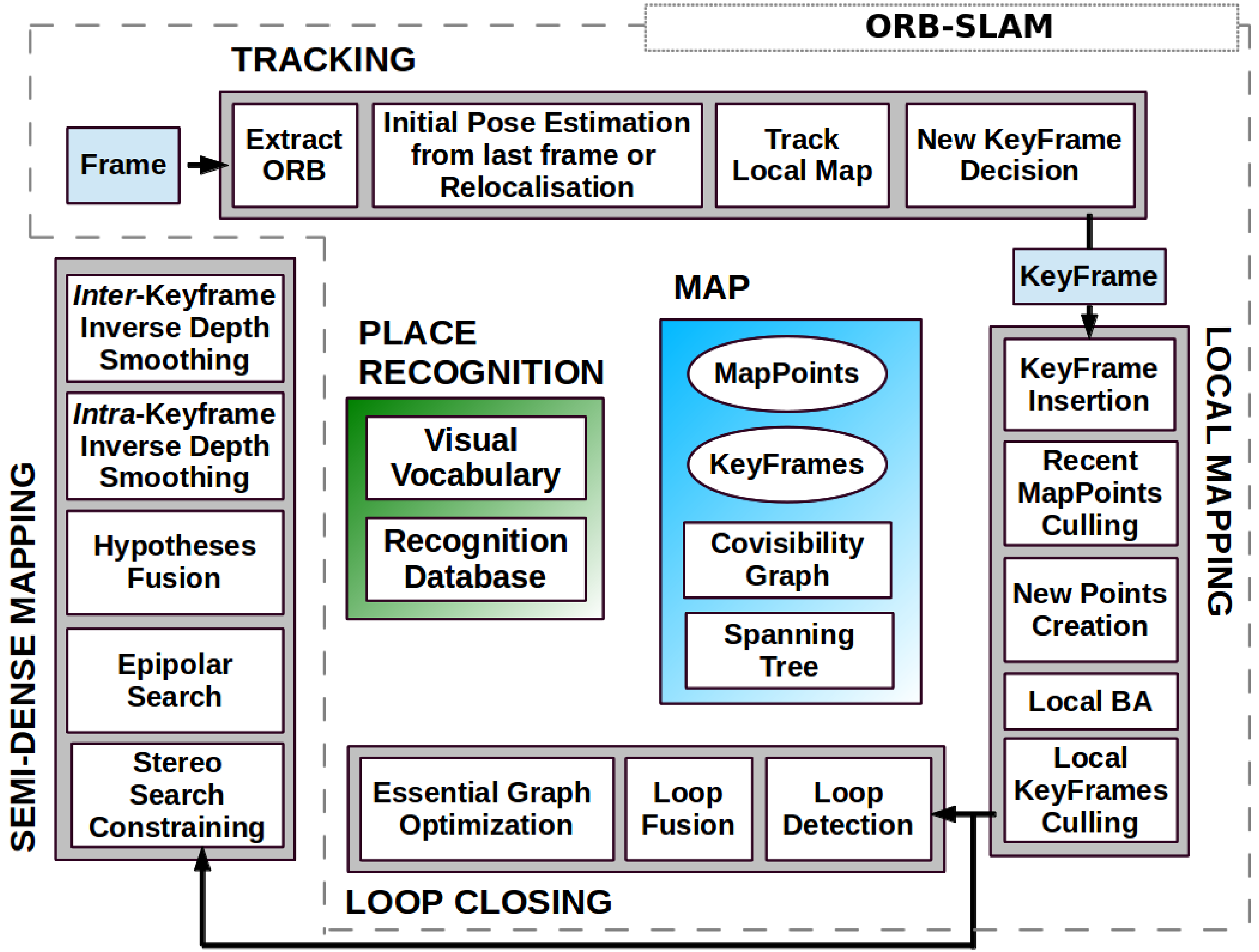
- **Bundle adjustment (joint map-trajectory optimization)**

Place Recogniton (loop detection, relocalization)

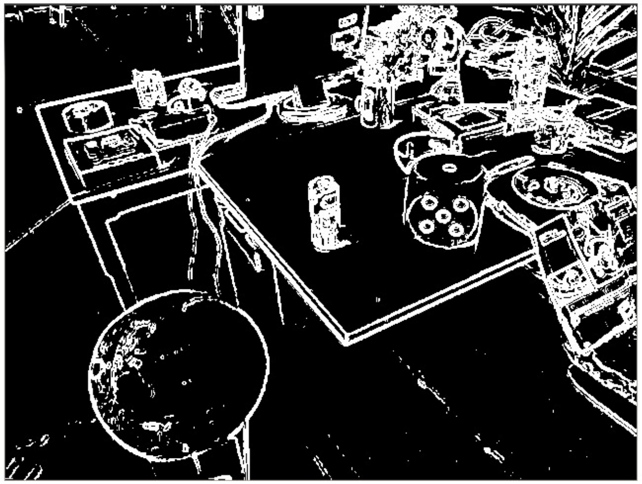
- **Bags of Words**

But sparse reconstructions ...

Probabilistic Semi-Dense Mapping



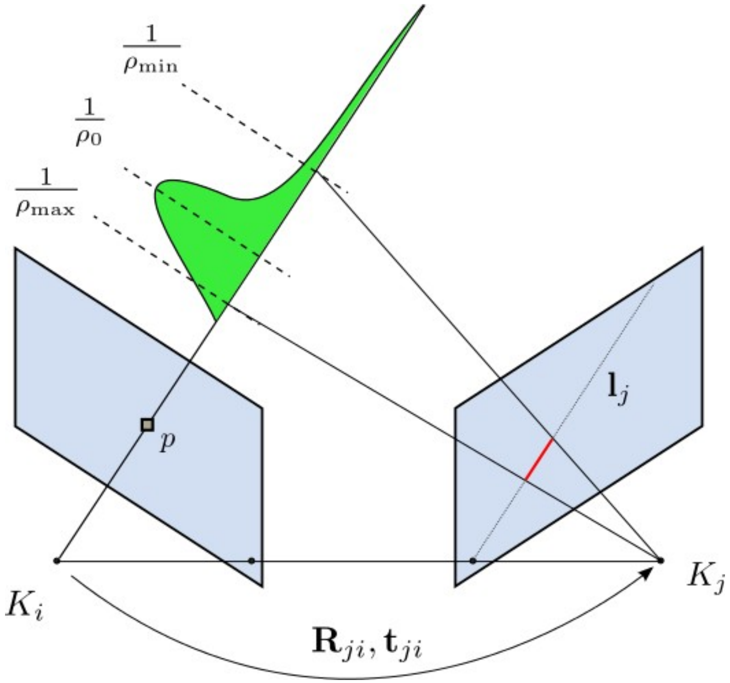
Probabilistic Semi-Dense Mapping

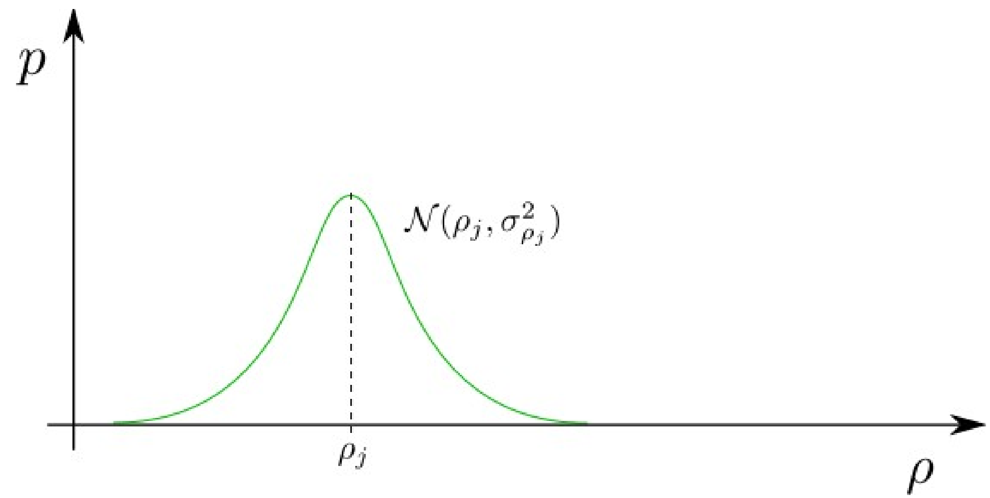
		?
KeyFrame	High Gradient Pixels	Inverse Depth Map & uncertainty

Compute each inverse depth map from scratch using neighbor keyframes

Probabilistic Semi-Dense Mapping

Per Pixel Operations





Epipolar search in neighbor keyframes

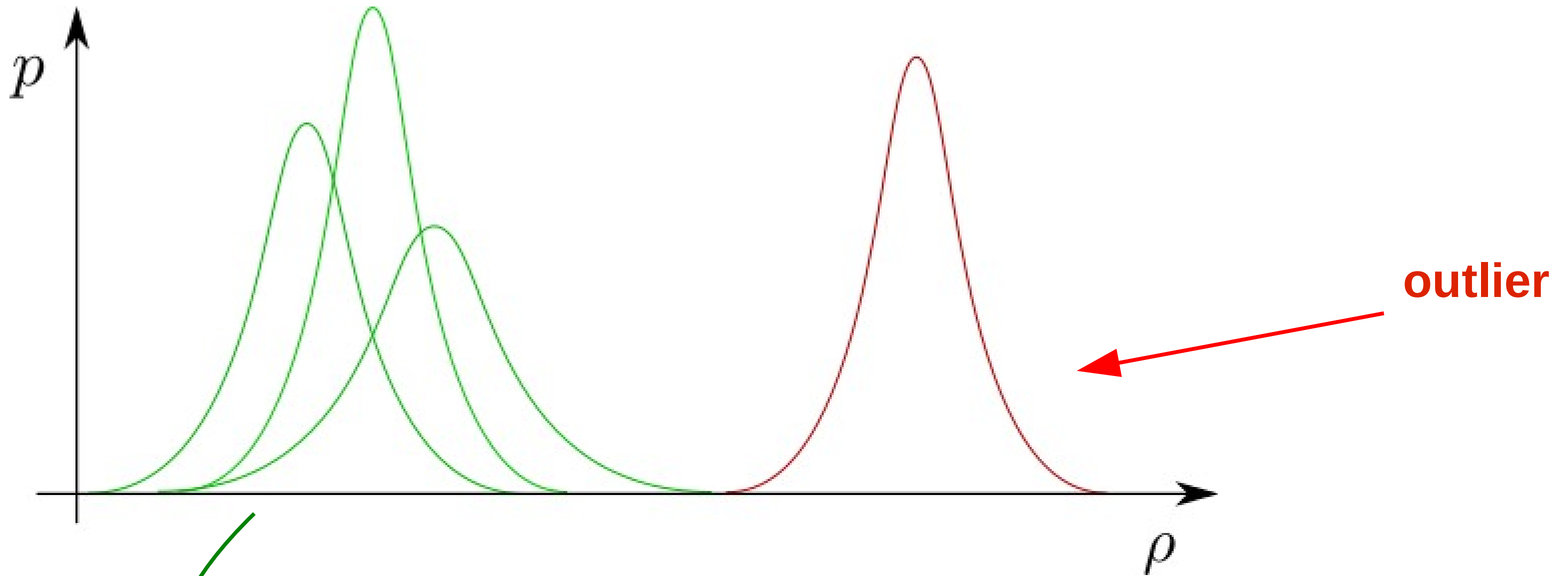
Generate an inverse depth hypothesis per neighbor keyframe

Compare { Intensity, Gradient Modulo, Gradient Direction }

Uncertainty { Image noise, Parallax, Matching ambiguity }

Probabilistic Semi-Dense Mapping

Per Pixel Operations

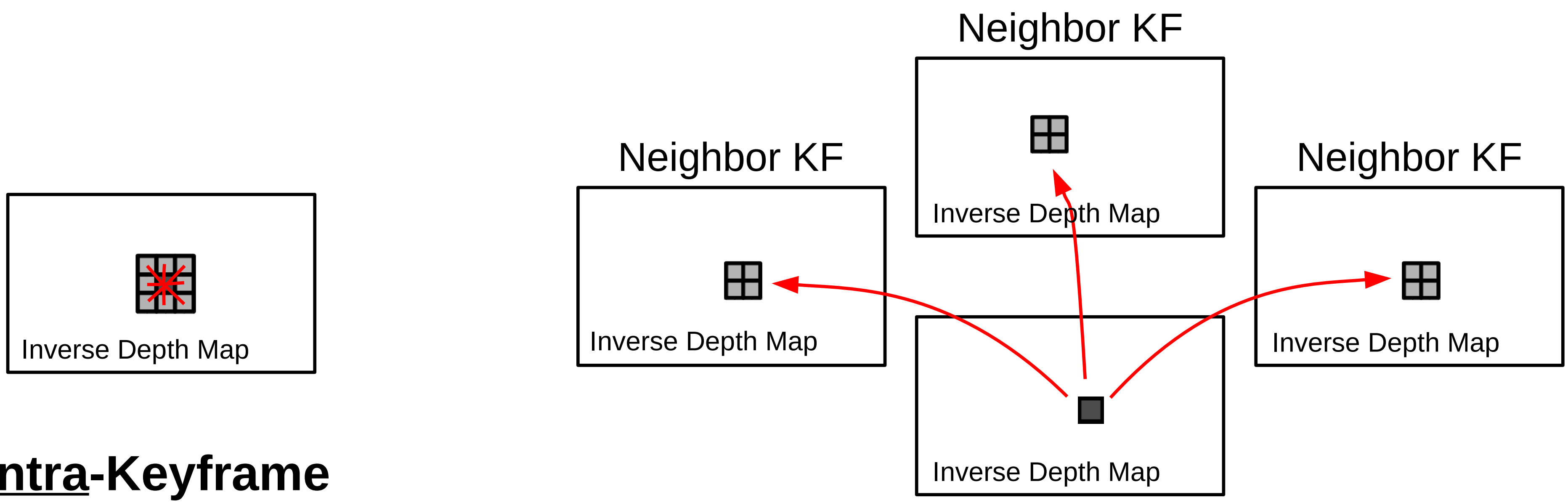


Consistent Hypotheses Fusion

$$\rho_p = \frac{\sum_j \frac{1}{\sigma^2_{\rho_j}} \rho_j}{\sum_j \frac{1}{\sigma^2_{\rho_j}}}, \quad \sigma^2_{\rho_p} = \frac{1}{\sum_j \frac{1}{\sigma^2_{\rho_j}}}$$

Probabilistic Semi-Dense Mapping

Per Pixel Operations



Intra-Keyframe

Inter-Keyframe

Smoothing

Outliers Detection

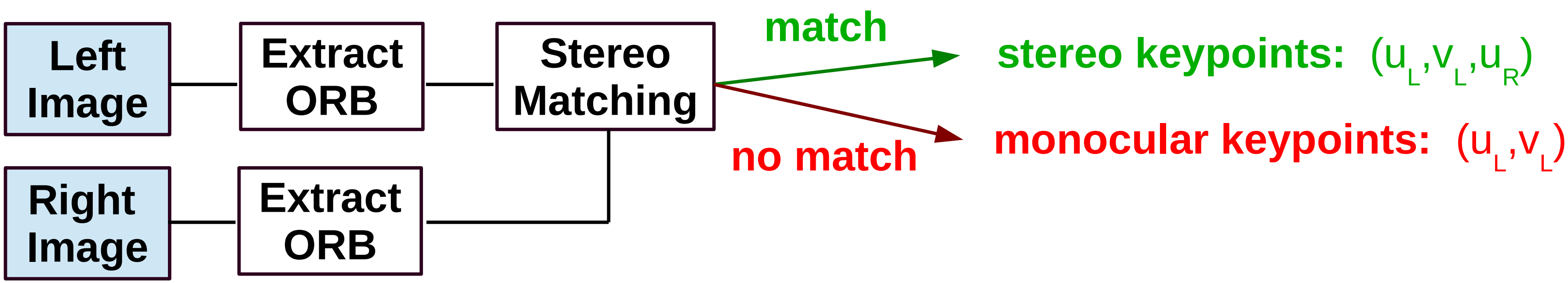
Probabilistic Semi-Dense Mapping

https://youtu.be/HlBmq70LKrQ

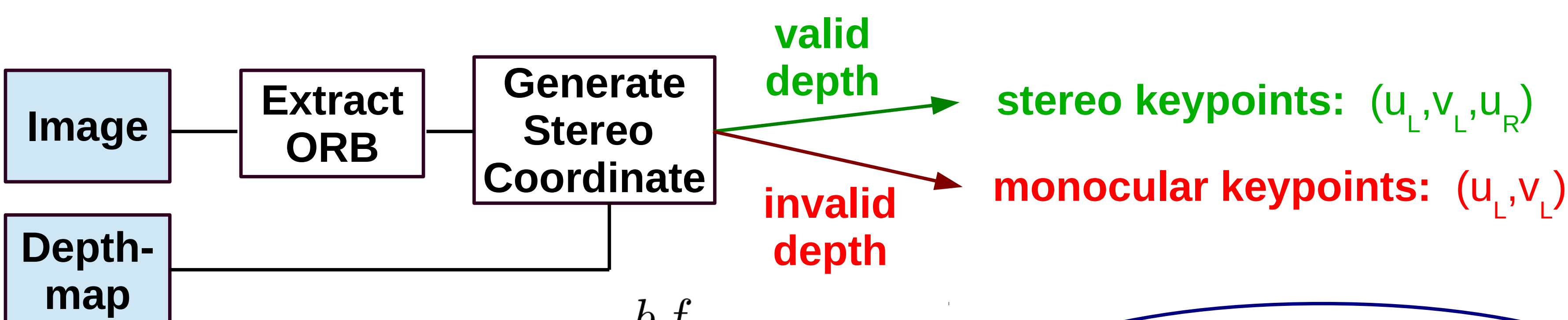
Stereo/RGBD ORB-SLAM

Universidad Zaragoza

Stereo (rectified)



RGBD



$$u_R = u_L - \frac{b\,f}{z}$$

Bundle Adjustment with monocular and stereo measurements

Stereo/RGBD ORB-SLAM

***Close stereo* keypoints** (depth < threshold)

Triangulation: 1 Frame

- ✓ **Fix scale**
- ✓ **Pure rotations**
- ✓ **Agility**

***Far stereo* keypoints** (depth > threshold)

Triangulation: 2 Keyframes

- ✓ **Moderate scale fix**
- ✓ **Improves rotation estimation**

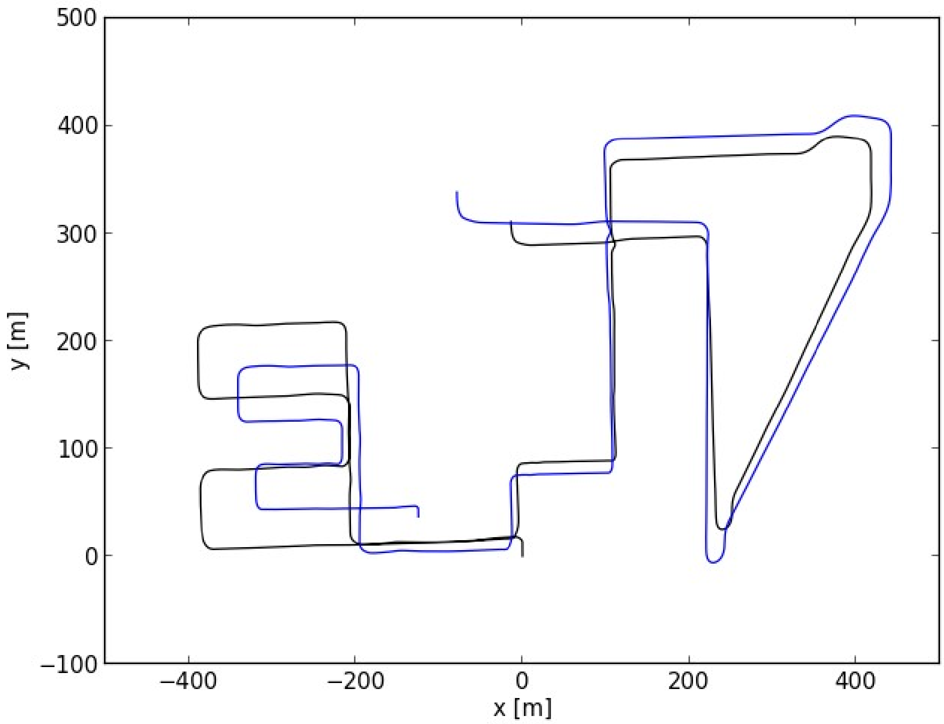
***Monocular* keypoints** (no stereo information)

Triangulation: 2 Keyframes

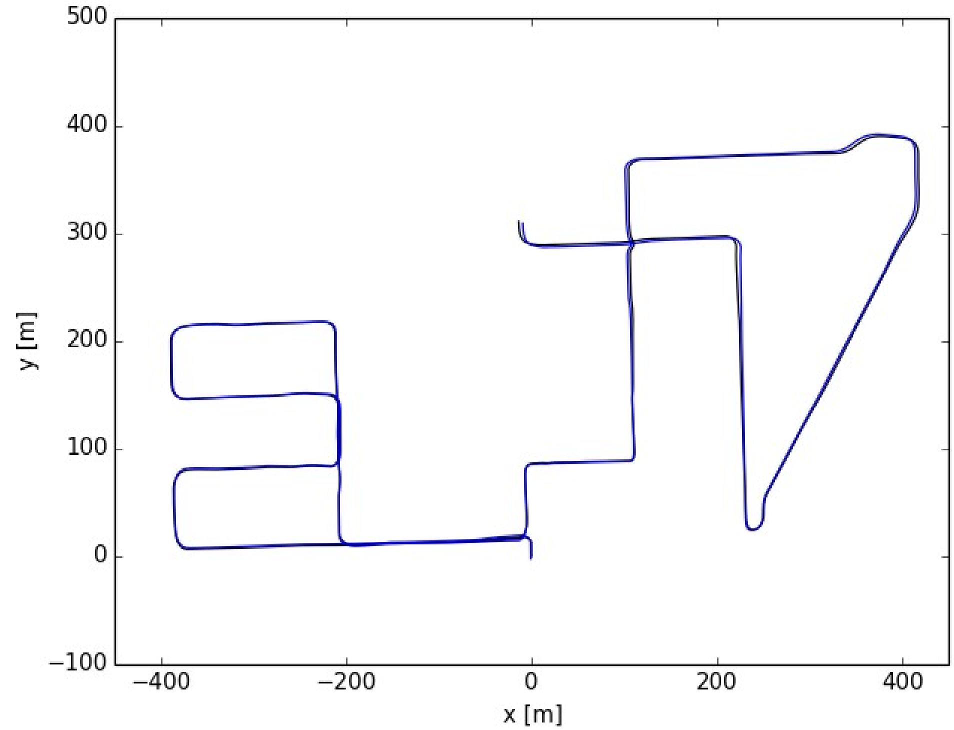
- ✓ **Robustness to sensor limitations (RGBD)**
- ✓ **Robustness to matching failure (stereo)**

Stereo/RGB-D ORB-SLAM

Monocular



Stereo



Scale drift!

Estimation
Ground Truth

Stereo ORB-SLAM

https://youtu.be/51NQvg5n-FE

Stereo/RGB-D ORB-SLAM

	Method	Setting	Code	Translation	Rotation	Runtime	Environment	Compare
1	V-LOAM			**0.75 %**	**0.0018 [deg/m]**	0.3 s	4 cores @ 2.5 Ghz (C/C++)	
J. Zhang and S. Singh: Visual-lidar Odometry and Mapping: Low- rift, Robust, and Fast. IEEE International Conference on Robotics and Automation(ICRA) 2015.								
2	LOAM			0.88 %	0.0022 [deg/m]	1.0 s	2 cores @ 2.5 Ghz (C/C++)	
J. Zhang and S. Singh: LOAM: Lidar Odometry and Mapping in Real-time. Robotics: Science and Systems Conference (RSS) 2014.								
3	ROCC			0.98 %	0.0028 [deg/m]	0.35 s	1 core @ 2.0 Ghz (C/C++)	
4	SOFT			1.03 %	0.0029 [deg/m]	0.1 s	2 cores @ 2.5 Ghz (C/C++)	
I. Cvišić and I. Petrović: Stereo odometry based on careful feature selection and tracking. European Conference on Mobile Robots (ECMR) 2015.								
5	cv4xv1-sc			1.09 %	0.0029 [deg/m]	0.145 s	GPU @ 3.5 Ghz (C/C++)	
M. Persson, T. Piccini, R. Mester and M. Felsberg: Robust Stereo Visual Odometry from Monocular Techniques. IEEE Intelligent Vehicles Symposium 2015.								
6	DEMO		code	1.14 %	0.0049 [deg/m]	0.1 s	4 cores @ 2.5 Ghz (C/C++)	
J. Zhang, M. Kaess and S. Singh: Real-time Depth Enhanced Monocular Odometry. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2014.								
7	S-ORB			1.20 %	0.0030 [deg/m]	0.06 s	2 cores @ >3.5 Ghz (C/C++)	
8	S-LSD-SLAM		code	1.20 %	0.0033 [deg/m]	0.07 s	1 core @ 3.5 Ghz (C/C++)	
J. Engel, J. St\"uckler and D. Cremers: Large-Scale Direct SLAM with Stereo Cameras. Int.~Conf.~on Intelligent Robot Systems (IROS) 2015.								
9	NOSM			1.21 %	0.0036 [deg/m]	0.8 s	2 cores @ 3.0 Ghz (C/C++)	
Anonymous submission								
10	FMVS			1.28 %	0.0023 [deg/m]	0.03 s	2 cores @ 2.5 Ghz (C/C++)	
11	MFI			1.30 %	0.0030 [deg/m]	0.1 s	1 core @ 2.2 Ghz (C/C++)	
H. Badino, A. Yamamoto and T. Kanade: Visual Odometry by Multi-frame Feature Integration. First International Workshop on Computer Vision for Autonomous Driving at ICCV 2013.								
12	S-PTAM		code	1.35 %	0.0023 [deg/m]	0.08 s	4 cores @ 2.2 Ghz (C/C++)	
T. Pire, T. Fischer, J. Civera, P. Cristóforis and J. Berlles: Stereo Parallel Tracking and Mapping for Robot localization. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2015.								
13	TLBBA			1.36 %	0.0038 [deg/m]	0.1 s	1 Core @2.8GHz (C/C++)	
W. Lu, Z. Xiang and J. Liu: High-performance visual odometry with two- stage local binocular BA and GPU. Intelligent Vehicles Symposium (IV), 2013 IEEE 2013.								
14	2FO-CC			1.37 %	0.0035 [deg/m]	0.1 s	1 core @ 3.0 Ghz (C/C++)	
I. Krešo and S. Šegvić: Improving the Egomotion Estimation by Correcting the Calibration Bias. VISAPP 2015.								
15	VoBa			1.46 %	0.0030 [deg/m]	0.1 s	1 core @ 2.0 Ghz (C/C++)	
16	TDVO			1.47 %	0.0030 [deg/m]	0.2 s	8 cores @ 3.5 Ghz (Matlab + C/C++)	
Anonymous submission								
17	JDO			1.50 %	0.0039 [deg/m]	0.06 s	1 core @ >3.5 Ghz (C/C++)	

Stereo ORB-SLAM

https://youtu.be/dF7_I2Lin54

RGBD ORB-SLAM

https://youtu.be/LnbAI-o7YHk

Universidad Zaragoza 1542

Monocular ORB-SLAM

R. Mur-Artal, J. M. M. Montiel and J. D. Tardos. *A versatile and Accurate Monocular SLAM System*. IEEE Transactions on Robotics. 2015

Open-source

github.com/raulmur/ORB_SLAM

Monocular Semi-Dense Mapping

R. Mur-Artal and J. D. Tardos. *Probabilistic Semi-Dense Mapping from Highly Accurate Feature-Based Monocular SLAM.* Robotics: Science and Systems. 2015

Stereo/RGBD ORB-SLAM

Unpublished

Open-source coming soon!

ROBOTICS UNIVERSIDAD DE ZARAGOZA

The Future of Real-Time SLAM
ICCV 2015 Workshop

Should we still do sparse-feature based SLAM?

Raúl Mur Artal
PhD. student, University of Zaragoza