Skoltech

Skolkovo Institute of Science and Technology

L16: Introduction to Visual SLAM

Perception in Robotics

Prof. Gonzalo Ferrer,

Skoltech, 16 March 2021

Mono SLAM (Davison 2007)

The first successful application of the SLAM methodology from mobile robotics to the “pure vision” domain of a single uncontrolled camera.

Idea: Online-SLAM using EKF with the following particularities:

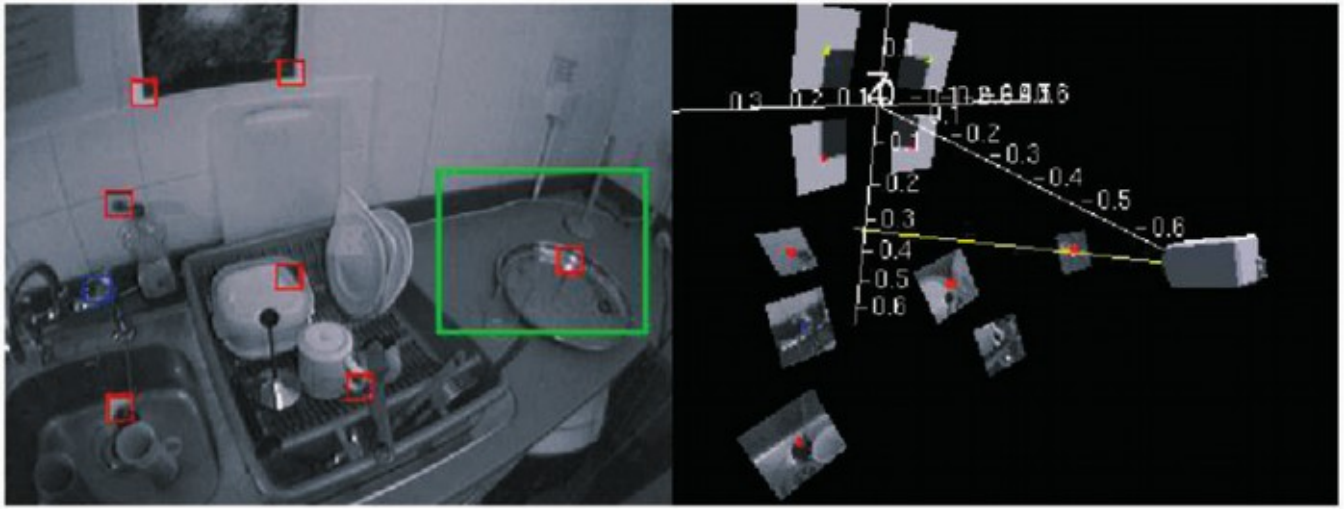
- Sparse but persistent features (image patches)
- General motion model for smooth camera movement.
- Real-time and drift-free
- Feature initialization and feature orientation estimation (this is going to be a sentivie issue for all visual SLAM algorithms).

Davison et al. “MonoSLAM: Real-Time Single Camera SLAM”, PAMI 2007

Mono SLAM (Davison 2007)

State variable: $y_t = [x_t, m_1, m_2, \ldots, m_N]$

The camera pose includes also velocities.



In this image we can see and example of MonoSLAM, on the left, the example of the planar patches (landmarks) and on the right, the estimation of its position and the camera pose.

Davison et al. “MonoSLAM: Real-Time Single Camera SLAM”, PAMI 2007

PTAM (Klein 2007)

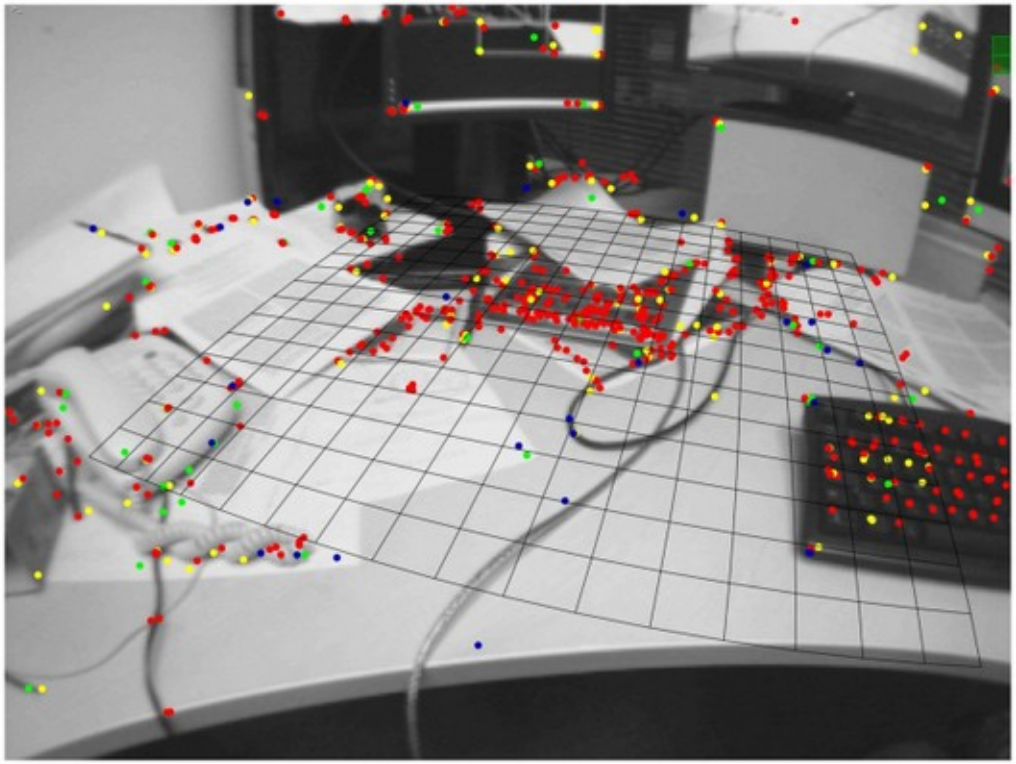
Parallel Tracking and Mapping (PTAM):

- Introducing the concept of Keyframe.
- Splits Tracking and mapping, running in two parallel threads.
- Camera Tracking (localization) is done by aligning the current image with the map. To do this, points in the map are projected to the image and matches are the camera pose is optimized.
- Mapping is calculated by graph optimization of keypoints in the keyframe.
- The map is densely initialized from a stereo pair.

https://www.youtube.com/watch?v=F3s3M0mokNc

Klein and Murray “Parallel Tracking and Mapping for Small AR Workspaces”, ISMAR 2007

PTAM (Klein 2007)



In this example, the on-line map contains around 3000 points, of which the system attempted to find 100 in the current frame.

Klein and Murray “Parallel Tracking and Mapping for Small AR Workspaces”, ISMAR 2007

Why Filtering? (Strasdat 2010)

Comparing MonoSLAM with PTAM for real-time applications.

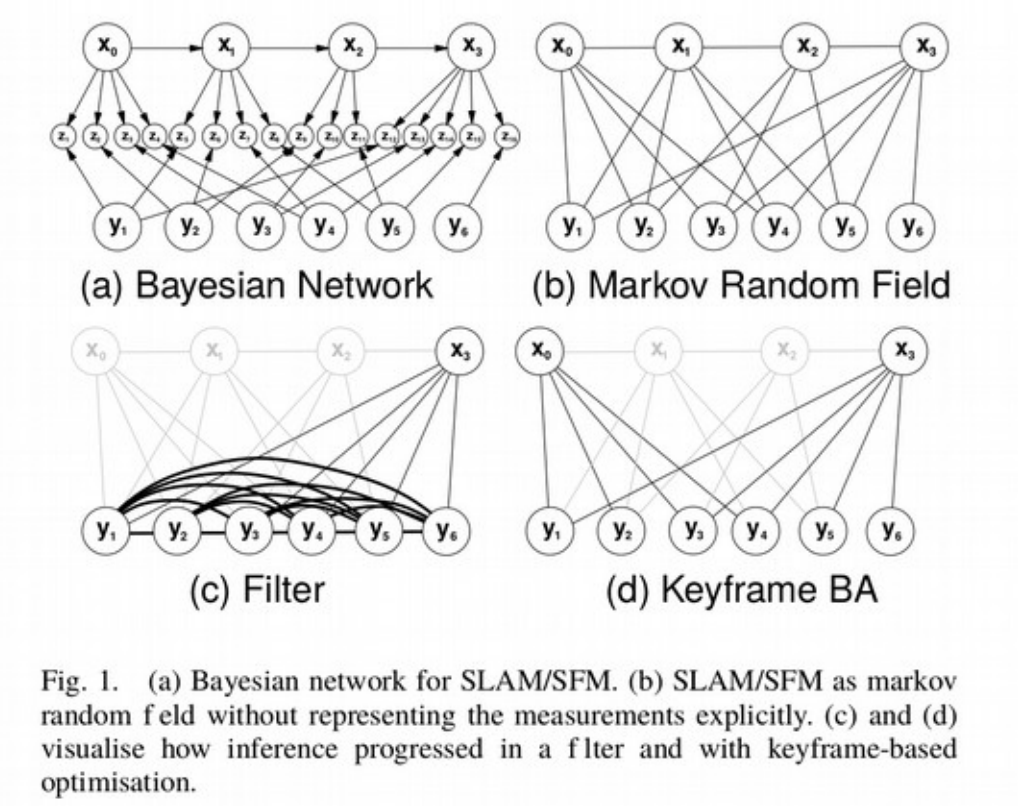


Fig. 1. (a) Bayesian network for SLAM/SFM. (b) SLAM/SFM as markov random f eld without representing the measurements explicitly. (c) and (d) visualise how inference progressed in a f lter and with keyframe-based optimisation.

Strasdat, Montiel and Davison “Real-time Monocular SLAM: Why Filter?”, ICRA 2010

Why Filtering? (Strasdat 2010)

After a fair comparison from both methods:

- Increasing the number of features improves the results, while increasing the number of intermediate frames has a minor impact
- Filtering is only superior for very small processing budgets and smoothing is superior otherwise.

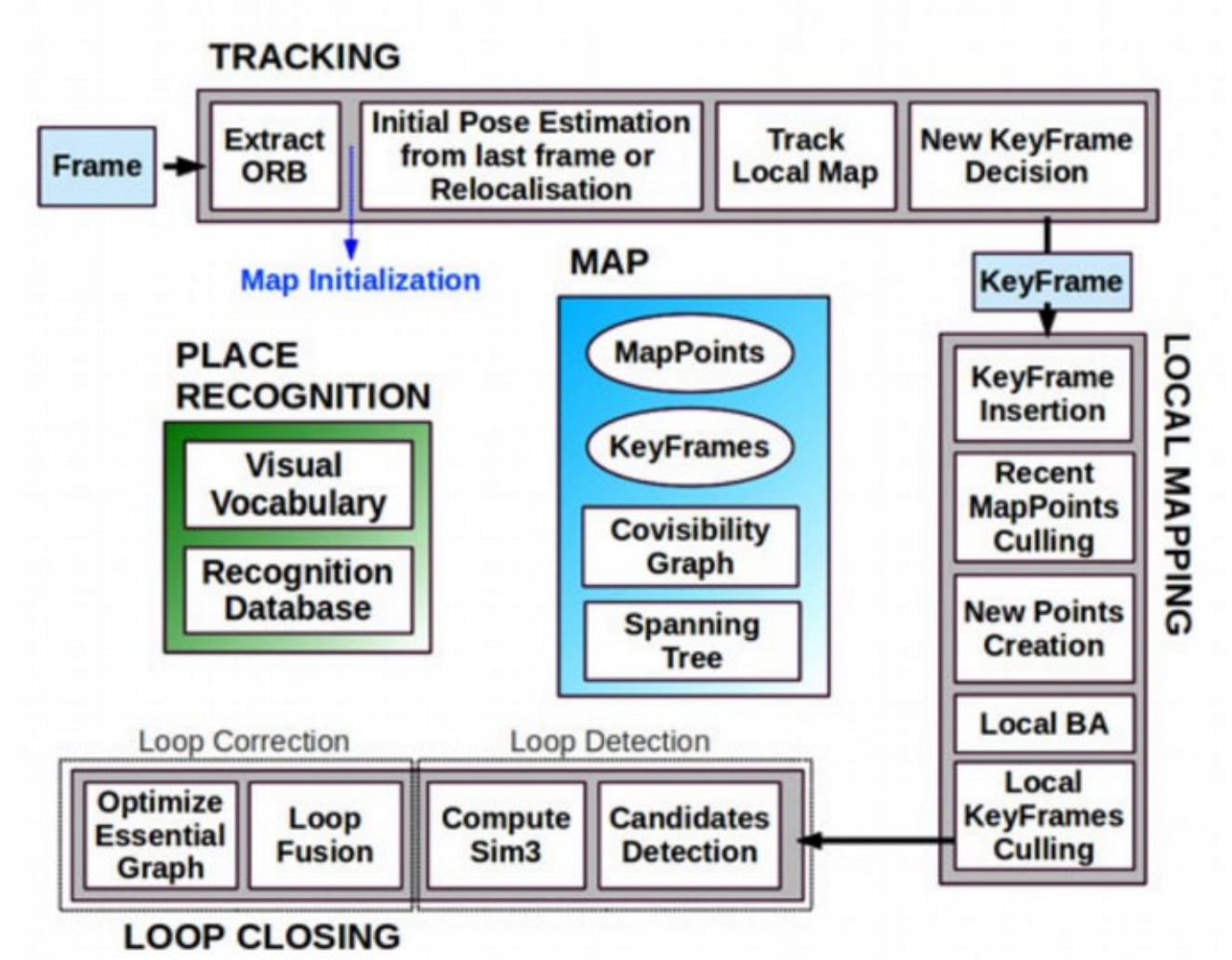
Strasdat, Montiel and Davison “Real-time Monocular SLAM: Why Filter?”, ICRA 2010

ORB-SLAM (Mur-Artal 2015)

- Current state of the art implementation of key-frame-based visual SLAM including all necessary modules.
- It uses the same keypoint descriptors for tracking, mapping, re-localization and loop closing.
- A survival of the fittest strategy to select landmarks and keyframes, allows a refinement over time.
- Flexible in the configuration of many number of parmeters/descriptors.
- Real-time on CPU.
- Open source and ready to use tool for visual SLAM. There are several follow ups for stereo/RGB-D (ORB-SLAM 2) and IMU (ORB-SLAM 3).

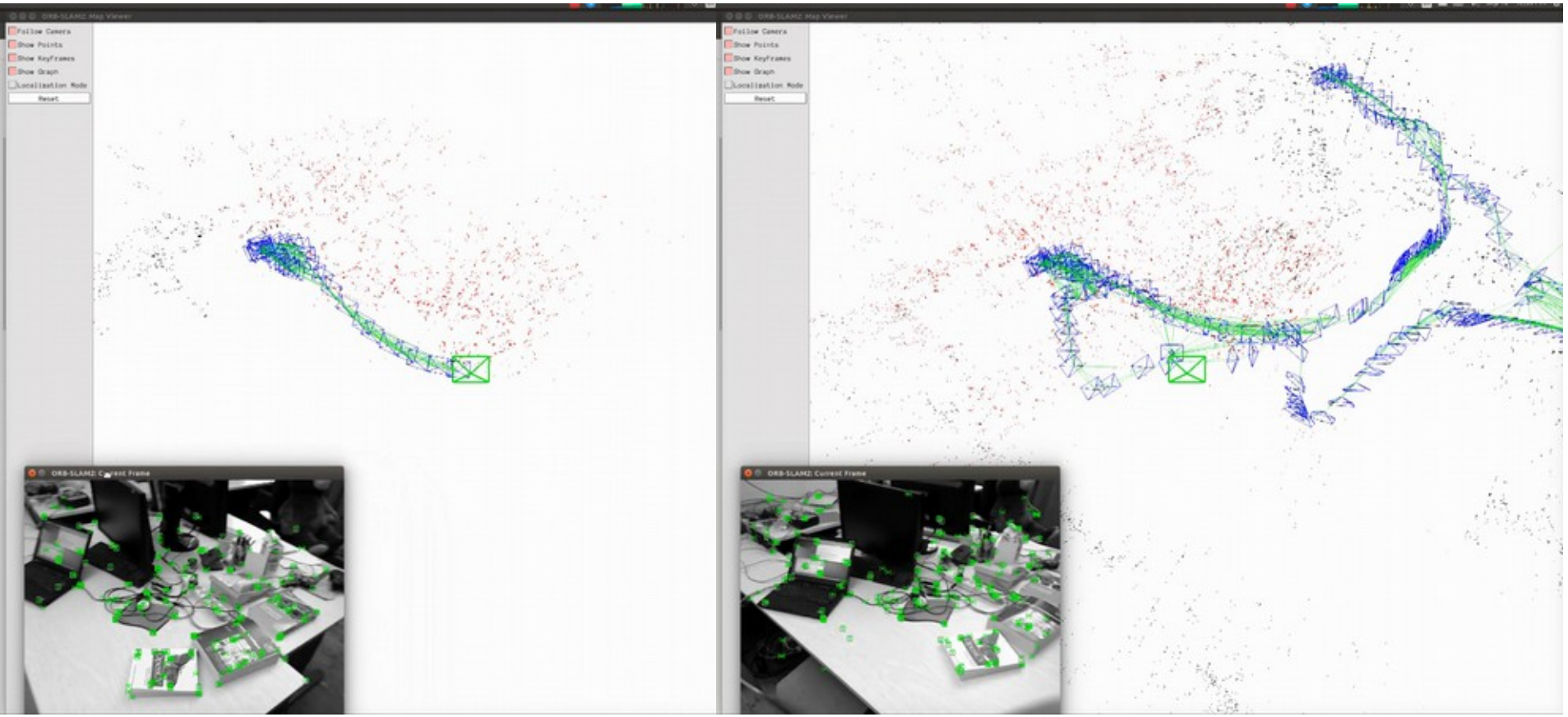
Mur-Artal, Montiel and Tardos “ORB-SLAM: A Versatile and Accurate Monocular SLAM System”, TRO 2015

ORB-SLAM (Mur-Artal 2015)



Mur-Artal, Montiel and Tardos “ORB-SLAM: A Versatile and Accurate Monocular SLAM System”, TRO 2015

ORB-SLAM (Mur-Artal 2015)



Mur-Artal, Montiel and Tardos “ORB-SLAM: A Versatile and Accurate Monocular SLAM System”, TRO 2015

DTAM (Newcombe 2011)

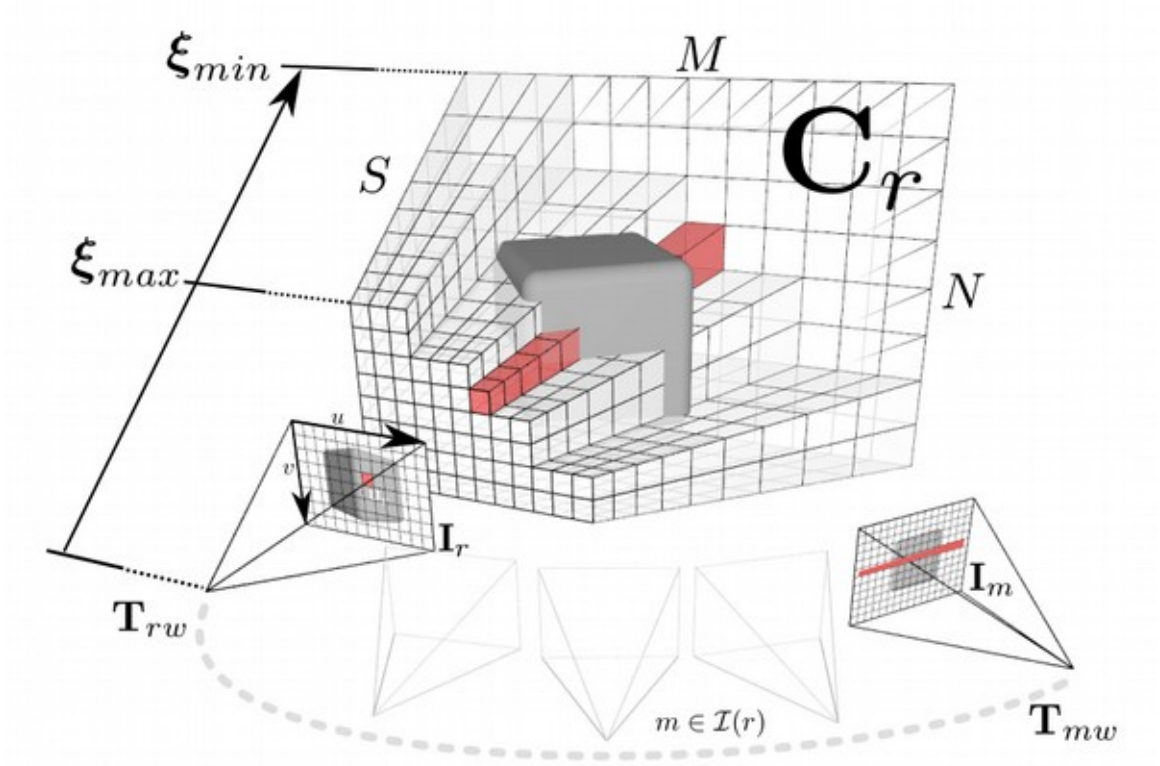
Dense Tracking and Mapping

- Not feature extraction (keypoints) but uses dense (every pixel) reconstruction.
- Volumetric reconstruction (grid) of each keyframe with millions of vertices.
- Camera motion estimation at frame-rate.
- Optimization is involved.

Why dense methods then? They use all data in the image to get more complete results. See Platinsky et al. “Monocular visual odometry: Sparse joint optimisation or dense alternation?”. ICRA 2017

Newcombe, Lovegrove and Davison “DTAM: Dense Tracking and Mapping in Real-Time”, ICCV 2011

DTAM (Newcombe 2011)



A keyframe consists of a reference image Ir and a data cost volume Cr. Each pixel has an associated row of entries (in red) that store the average photometric error.

Newcombe, Lovegrove and Davison “DTAM: Dense Tracking and Mapping in Real-Time”, ICCV 2011

DTAM (Newcombe 2011)

A sequence of images is required.

Tracking: Given a dense model (keyframe) we can align it with our current image observation by projecting the volume Cr into the image plane and minimizing the photometric error between the observation (image) and the keyframe (Volume Cr).

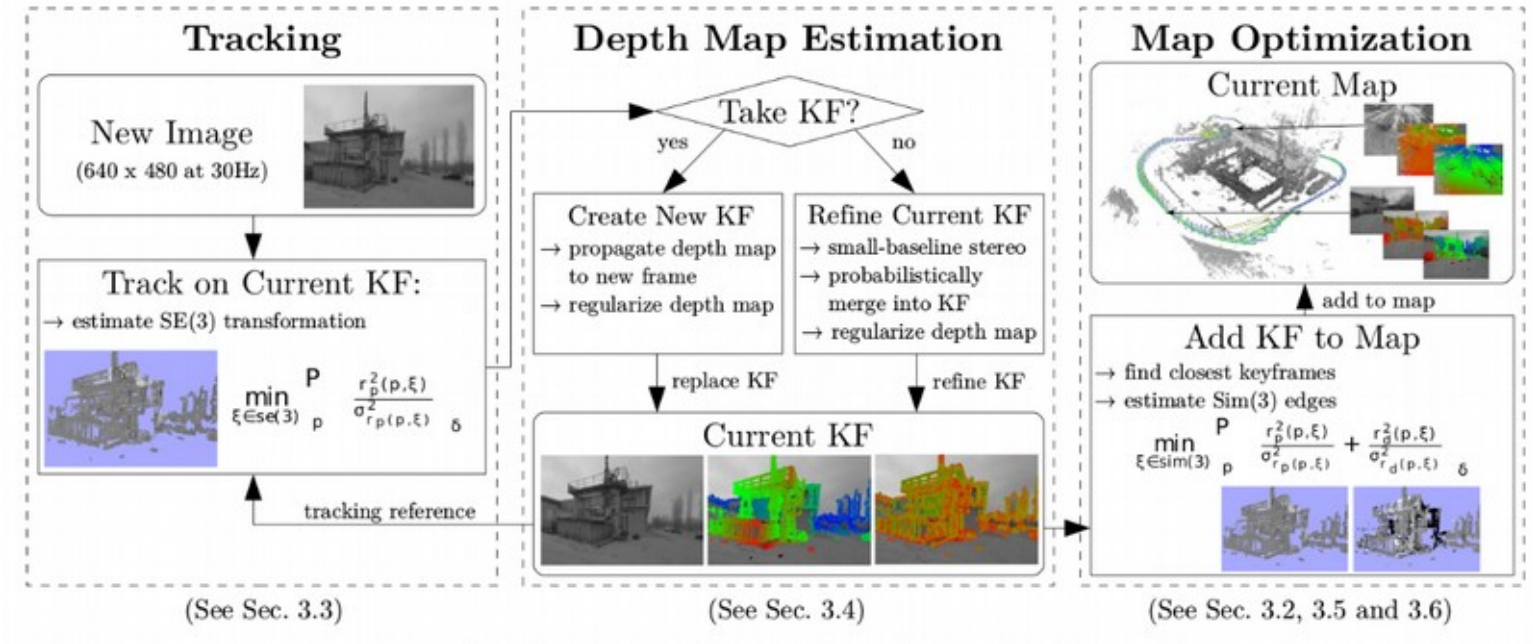
Newcombe, Lovegrove and Davison “DTAM: Dense Tracking and Mapping in Real-Time”, ICCV 2011

LSD-SLAM (Engel 2014)

Large-Scale Direct Monocular SLAM:

- Semi-dense depth reconstructions: Only those pixels with non-zero gradient are considered.
- Large-scale maps.
- Scale drift and scale ambiguity solved by using sim(3), instead of SE(3).
- Global map as a pose graph consisting of keyframes.

LSD-SLAM (Engel 2014)



- Tracking of new camera poses w.r.t keyframe.
- Depth maps estimation, refines current keyframe or creates a new one.
- Map optimization (SLAM, Bundle Adjustment, Structure from Motion)

Engel, Schops and Cremers “LSD-SLAM: Large-Scale Direct Monocular SLAM”, ECCV 2014

OKVIS (Leutenegger 2015)

Keyframe-Based Visual-Intertial SLAM Using Nonlinear Optimization:

Windowed-based batch optimization of images and IMU measurements.

Sensor fusion of observations in a **tight** optimization of both camera and inertial measurements.

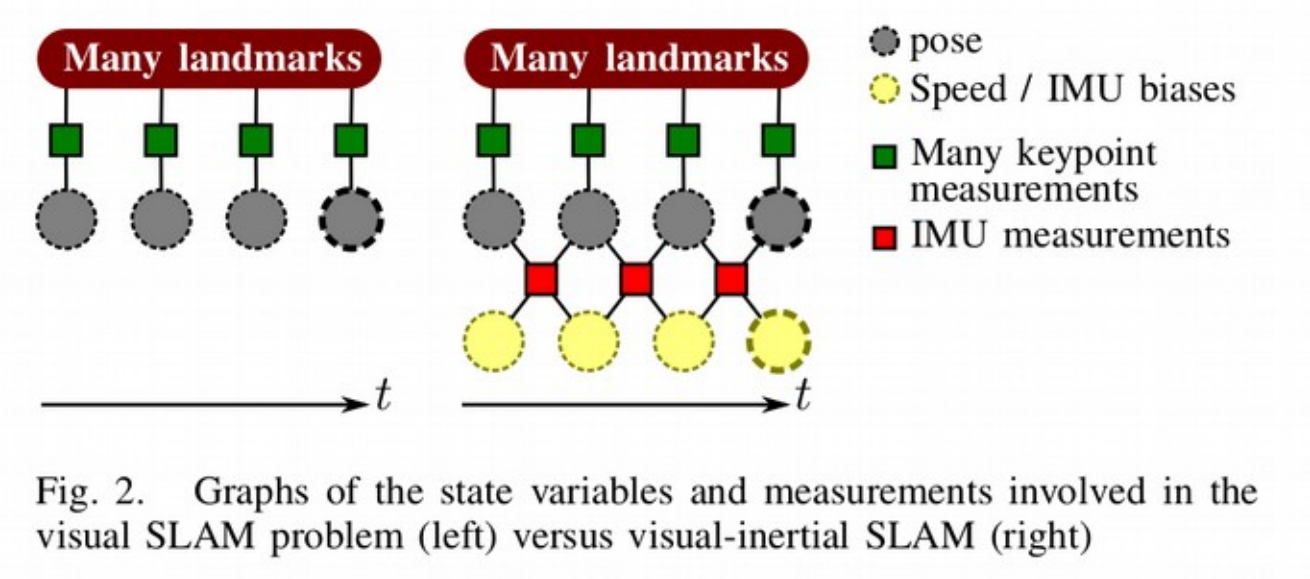


Fig. 2. Graphs of the state variables and measurements involved in the visual SLAM problem (left) versus visual-inertial SLAM (right)

Leutenegger et al. “Keyframe-Based Visual-Intertial SLAM Using Nonlinear Optimization”, IJRR 2015