About SLAM

SLAM stands for simultaneous localisation and mapping and is a concept which solves part of the navigation problem of autonomous mobile robotics, like self-driving vehicles. It is made up of two parts:

- Mapping - building a map of the environment in which the robot operates localisation.
- Navigating this environment using the map while keeping track of the robot's relative position and orientation.

Large-Scale Direct monocular SLAM Simultaneous Localisation and Mapping (LSD-SLAM) method which not only locally tracks the motion of the camera, but allows to build consistent, large-scale maps of the environment. It uses direct image alignment coupled with filtering-based estimation of semi-dense depth maps. The global map is represented as a pose graph consisting of keyframes as vertices with 3D similarity transforms as edges, elegantly incorporating changing scale of the environment and allowing to detect and correct accumulated drift. The method runs in real-time on a CPU.

Many variants available:

- LSD-SLAM with Stereo Cameras
- LSD-SLAM for Omnidirectional Cameras
- Monocular LSD-SLAM

LSD-SLAM with stereo cameras

LSD-SLAM is a key-frame based localization and mapping approach which uses the following main steps:

- The motion of the camera is tracked towards a reference keyframe in the map. New keyframes are generated if the camera moved too far from existing keyframes in the map.

- Depth in the current reference keyframe is estimated from stereo correspondences based on the tracked mo- tion (temporal stereo)

- The poses of the keyframes are made globally consis- tent by mutual direct image alignment and pose graph optimization.

In Stereo LSD-SLAM, the depth in keyframes is in addition directly estimated from static stereo (see Fig. 2). There is a number of advantages of this approach to relying solely on temporal or solely on static stereo. Static stereo allows for estimating the absolute scale of the world and is independent of the camera movement. However, static stereo is constrained to a constant baseline (with, in many cases, a fixed direction), which effectively limits the performance to a specific range. Temporal stereo does not limit the perfor- mance to a specific range as demonstrated. The same sensor can be used in very small and very large

environments, and seamlessly transits between the two. On the other hand, it does not provide scale and requires non-degenerate camera movement. An additional benefit of combining temporal and static stereo is, that multiple baseline directions are available: while static stereo typically has a horizontal baseline – which does not allow for estimating depth along horizontal edges, temporal stereo allows for completing the depth map by providing other motion directions.

- In detail, we make the following key contributions:
  - We generalize LSD-SLAM to stereo cameras, combining temporal and static stereo in a direct, real-time capable SLAM method.
  - We explicitly model illumination changes during direct image alignment, thereby making the method highly robust even in challenging real-world conditions.
  - We perform a systematic evaluation on two benchmark datasets from realistic robotics applications, demonstrating the state-of-the-art performance of our approach.

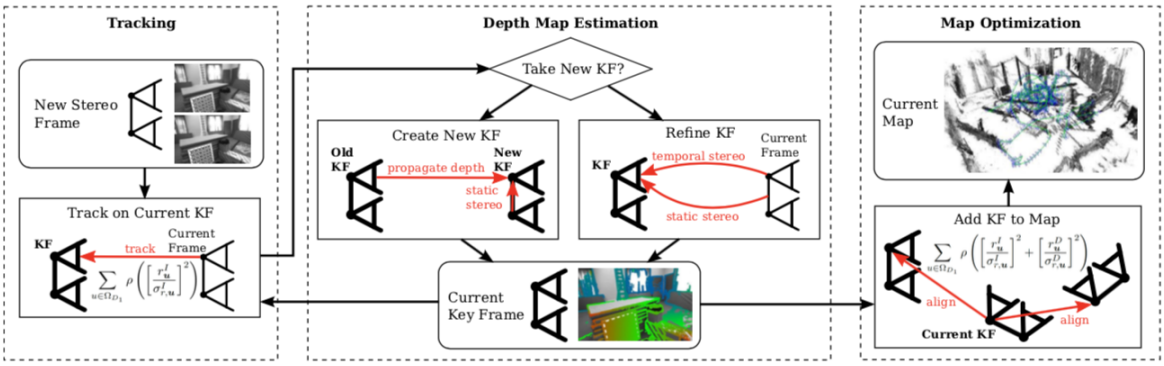
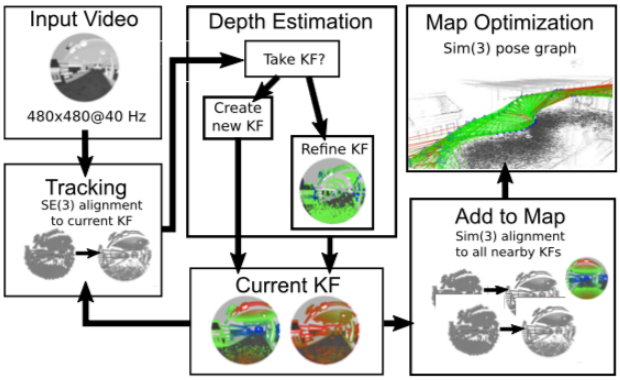


Fig. 2. Overview on the Stereo LSD-SLAM system.

LSD-SLAM for Omnidirectional Cameras

We propose a real-time, direct monocular SLAM method for omnidirectional or wide field-of-view fisheye cameras. Both tracking (direct image alignment) and mapping (pixel-wise distance filtering) are directly formulated for the unified omnidirectional model, which can model central imaging devices with a field of view well above 150◦. This is in stark contrast to existing direct mono-SLAM approaches like DTAM or LSD-SLAM, which operate on rectified images, limiting the field of view to well below 180◦. Not only does this allow to observe – and reconstruct – a larger portion of the surrounding environment, but it also makes the system more robust to degenerate (rotation-only) movement. The two main contribution are (1) the formulation of direct image alignment for the unified omnidirectional model, and (2) a fast yet accurate approach to incremental stereo directly on distorted images. We evaluated our framework on real-world sequences taken with a 185◦ fish-eye lens, and compare it to a rectified and a piecewise rectified approach.



LSD-SLAM Monocular

One of the major benefits of monocular SLAM – and simultaneously one of the biggest challenges – comes with the inherent scale-ambiguity: The scale of the world cannot be observed and drifts over time, being one of the major error sources. The advantage is that this allows to seamlessly switch between differently scaled environments, such as a desk environment indoors and large-scale outdoor environments. Scaled sensors on the other hand, such as depth or stereo cameras, have a limited range at which they can provide reliable measurements and hence do not provide this flexibility.

The algorithm consists of three major components: **tracking, depth map estimation and map optimization** as visualized in Fig. 3:

– The **tracking** component continuously tracks new camera images. That is, it estimates their rigid body pose $\xi \in se(3)$ with respect to the current keyframe, using the pose of the previous frame as initialization.

– The **depth map estimation** component uses tracked frames to either refine or replace the current keyframe. Depth is refined by filtering over many per-pixel, small-baseline stereo comparisons coupled with interleaved spatial regularization as originally proposed in [9]. If the camera has moved too far, a new keyframe is initialized by projecting points from existing, close-by keyframes into it.

– Once a keyframe is replaced as tracking reference – and hence its depth map will not be refined further – it is incorporated into the global map by the **map optimization** component. To detect loop closures and scale-drift, a similarity transform $\xi \in sim(3)$ to close-by existing keyframes (including its direct predecessor) is estimated using scale-aware, direct sim(3)-image alignment.

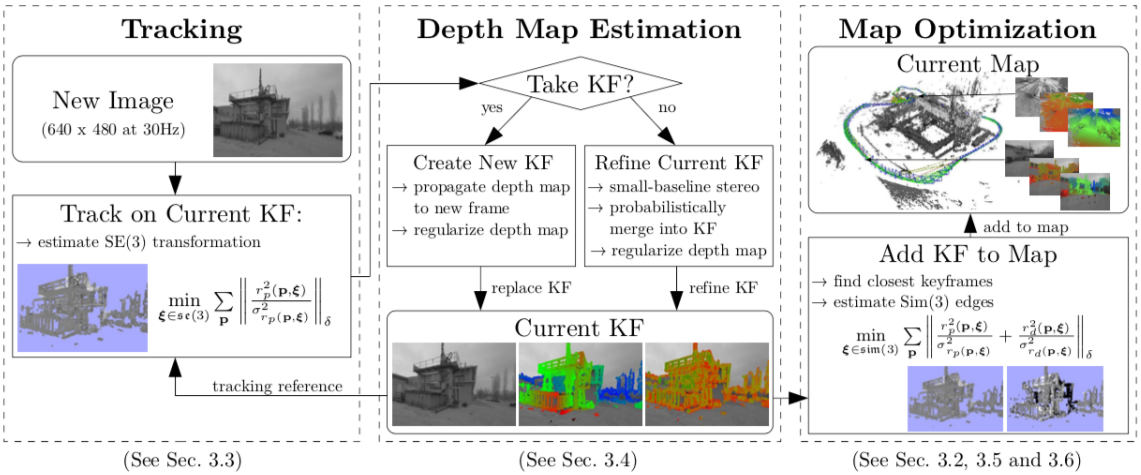


Fig. 3: Overview over the complete LSD-SLAM algorithm.

Initialization. To bootstrap the LSD-SLAM system, it is sufficient to initialize a first keyframe with a random depth map and large variance. Given sufficient translational camera movement in the first seconds, the algorithm "locks" to a certain configuration, and after a couple of keyframe propagations converges to a correct depth configuration. Some examples are shown in the attached video. A more thorough evaluation of this ability to converge without dedicated initial bootstrapping is outside the scope of this paper, and remains for future work.

General Papers:

- LSD-SLAM: Large-Scale Direct Monocular SLAM - Computer Vision …
- LSD-SLAM: Large-Scale Direct Monocular SLAM | SpringerLink
- LSD-SLAM: Large-Scale Direct Monocular SLAM (ECCV '14) - YouTube
- Monocular SLAM | LSD-SLAM
- GitHub - tum-vision/lsd_slam: LSD-SLAM
- Combining Feature-based and Direct Methods for ... - ais.uni-bonn.de