Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Chapter 8
Direct Approaches to Visual SLAM

Multiple View Geometry
Summer 2019

Prof. Daniel Cremers
Chair for Computer Vision and Artificial Intelligence
Departments of Informatics & Mathematics
Technical University of Munich

Overview

1. **Direct Methods**
2. **Realtime Dense Geometry**
3. **Dense RGB-D Tracking**
4. **Loop Closure and Global Consistency**
5. **Dense Tracking and Mapping**
6. **Large Scale Direct Monocular SLAM**
7. **Direct Sparse Odometry**

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Classical Approaches to Multiple View Reconstruction

In the past chapters we have studied classical approaches to multiple view reconstruction. These methods tackle the problem of structure and motion estimation (or visual SLAM) in several steps:

1. A set of feature points is extracted from the images – ideally points such as corners which can be reliably identified in subsequent images as well.
2. One determines a correspondence of these points across the various images. This can be done either through local tracking (using optical flow approaches) or by random sampling of possible partners based on a feature descriptor (SIFT, SURF, etc.) associated with each point.
3. The camera motion is estimated based on a set of corresponding points. In many approaches this is done by a series of algorithms such as the eight-point algorithm or the five-point algorithm followed by bundle adjustment.
4. For a given camera motion one can then compute a dense reconstruction using stereo reconstruction methods.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Shortcomings of Classical Approaches

Such classical approaches are indirect in the sense that they do not compute structure and motion directly from the images but rather from a sparse set of precomputed feature points. Despite a number of successes, they have several drawbacks:

- From the point of view of statistical inference, they are suboptimal: In the selection of feature points much potentially valuable information contained in the colors of each image is discarded.
- They invariably lack robustness: Errors in the point correspondence may have devastating effects on the estimated camera motion. Since one often selects very few point pairs only (8 points for the eight-point algorithm, 5 points for the five-point algorithm), any incorrect correspondence will lead to an incorrect motion estimate.
- They do not address the highly coupled problems of motion estimation and dense structure estimation. They merely do so for a sparse set of points. As a consequence, improvements in the estimated dense geometry will not be used to improve the camera motion estimates.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Toward Direct Approaches to Multiview Reconstruction

In the last few years, researchers have been promoting direct approaches to multi-view reconstruction. Rather than extracting a sparse set of feature points to determine the camera motion, direct methods aim at estimating camera motion and dense or semi-dense scene geometry directly from the input images. This has several advantages:

- Direct methods tend to be more robust to noise and other nuisances because they exploit all available input information.
- Direct methods provide a semi-dense geometric reconstruction of the scene which goes well beyond the sparse point cloud generated by the eight-point algorithm or bundle adjustment. Depending on the application, a separate dense reconstruction step may no longer be necessary.
- Direct methods are typically faster because the feature-point extraction and correspondence finding is omitted: They can provide fairly accurate camera motion and scene structure in real-time on a CPU.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

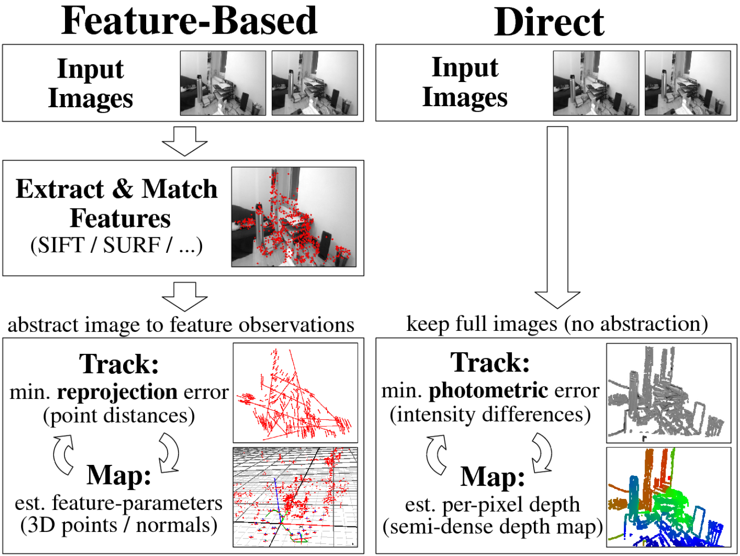
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Feature-Based versus Direct Methods



Engel, Sturm, Cremers, ICCV 2013

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Direct Methods for Multi-view Reconstruction

In the following, we will briefly review several recent works on direct methods for realtime multiple-view reconstruction:

- the method of Stühmer, Gumhold, Cremers, DAGM 2010 computes dense geometry from a handheld camera in real-time.
- the methods of Steinbrücker, Sturm, Cremers, 2011 and Kerl, Sturm, Cremers, 2013 directly compute the camera motion of an RGB-D camera.
- the method of Newcombe, Lovegrove, Davison, ICCV 2011 directly determines dense geometry and camera motion from the images.
- the method of Engel, Sturm, Cremers, ICCV 2013 and Engel, Schöps, Cremers, ECCV 2014 directly computes camera motion and semi-dense geometry for a handheld (monocular) camera.
- the method of Engel, Koltun, Cremers, PAMI 2018 directly estimates highly accurate camera motion and sparse geometry.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Realtime Dense Geometry from a Handheld Camera

Let $g_i \in SE(3)$ be the rigid body motion from the first camera to the i-th camera, and let $I_i : \Omega \rightarrow \mathbb{R}$ be the i-th image. A dense depth map $h : \Omega \rightarrow \mathbb{R}$ can be computed by solving the optimization problem:

$$\min_h \sum_{i=2}^{n} \int_\Omega \left| I_1(x) - I_i\big(\pi g_i(hx)\big) \right| \mathrm{d}x \; + \; \lambda \int_\Omega |\nabla h| \, \mathrm{d}x,$$

where x is represented in homogeneous coordinates and hx is the corresponding 3D point.

Like in optical flow estimation, the unknown depth map should be such that for all pixels $x \in \Omega$, the transformation into the other images I_i should give rise to the same color as in the reference image I_1.

This cost function can be minimized at framerate by coarse-to-fine linearization solved in parallel on a GPU.

Stuehmer, Gumhold, Cremers, DAGM 2010.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

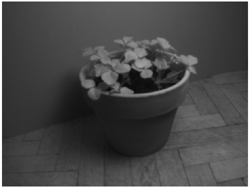
Loop Closure and Global Consistency

Dense Tracking and Mapping

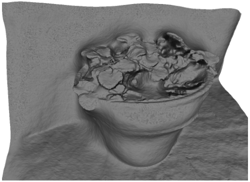
Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Realtime Dense Geometry from a Handheld Camera

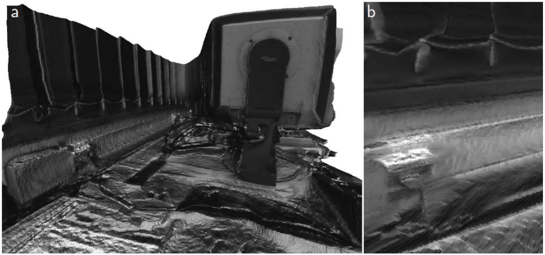


Input image

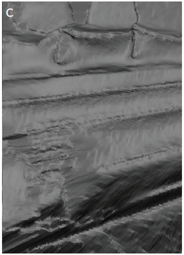


Reconstruction

Textured geometry



Textured reconstructions



Untextured

Stuehmer, Gumhold, Cremers, DAGM 2010.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

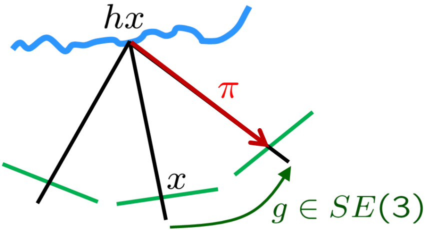
Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Dense RGB-D Tracking

The approach of Stühmer et al. (2010) relies on a sparse feature-point based camera tracker (PTAM) and computes dense geometry directly on the images. Steinbrücker, Sturm, Cremers (2011) propose a complementary approach to directly compute the camera motion from RGB-D images. The idea is to compute the rigid body motion g_ξ which optimally aligns two subsequent color images I_1 and I_2:

$$\min_{\xi \in \mathfrak{se}(3)} \int_\Omega \left| I_1(x) - I_2\left(\pi g_\xi(hx)\right)\right|^2 \mathrm{d}x$$



Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Dense RGB-D Tracking

The above non-convex problem can be approximated as a convex problem by linearizing the residuum around an initial guess ξ_0:

$$E(\xi) \approx \int_\Omega \left| I_1(x) - I_2\left(\pi g_{\xi_0}(hx)\right) - \nabla I_2^\top \left(\frac{d\pi}{dg_\xi}\right)\left(\frac{dg_\xi}{d\xi}\right)\xi \right|^2 dx$$

This is a convex quadratic cost function which gives rise to a linear optimality condition:

$$\frac{dE(\xi)}{d\xi} = A\xi + b = 0$$

To account for larger motions of the camera, this problem is solved in a coarse-to-fine manner. The linearization of the residuum is identical with a Gauss-Newton approach. It corresponds to an approximation of the Hessian by a positive definite matrix.

Steinbrücker, Sturm, Cremers 2011

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

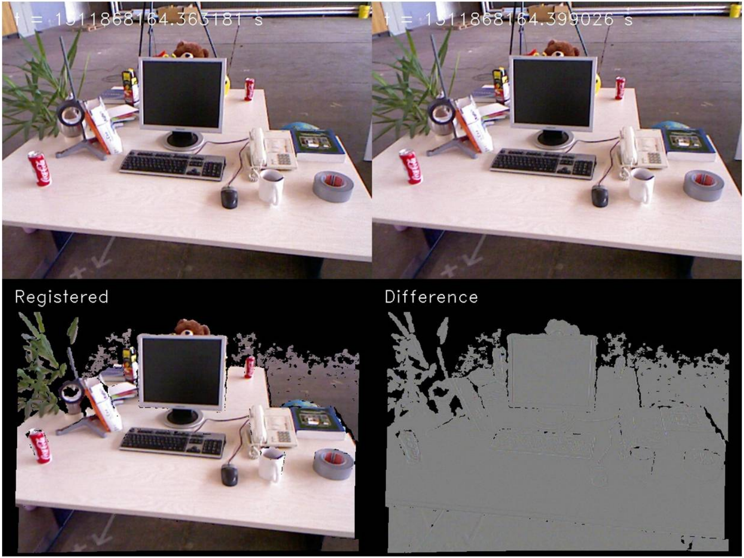
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Dense RGB-D Tracking



Steinbrücker, Sturm, Cremers 2011

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

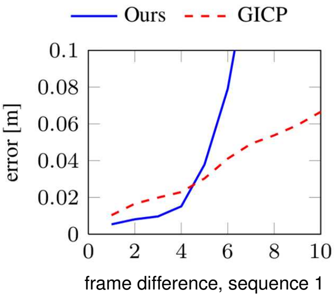
Dense Tracking and Mapping

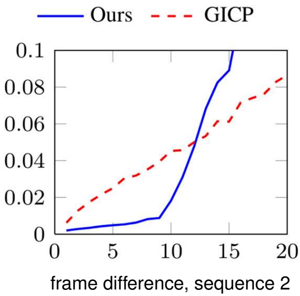
Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Dense RGB-D Tracking

In the small-baseline setting, this image aligning approach provides more accurate camera motion than the commonly used generalized Iterated Closest Points (GICP) approach.





Steinbrücker, Sturm, Cremers 2011

A related direct tracking approach was proposed for stereo reconstruction in Comport, Malis, Rives, ICRA 2007. A generalization which makes use of non-quadratic penalizers was proposed in Kerl, Sturm, Cremers, ICRA 2013.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

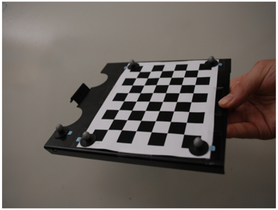
Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

A Benchmark for RGB-D Tracking

Accurately tracking the camera is among the most central challenges in computer vision. Quantitative performance of algorithms can be validated on benchmarks.



Sturm, Engelhard, Endres, Burgard, Cremers, IROS 2012

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

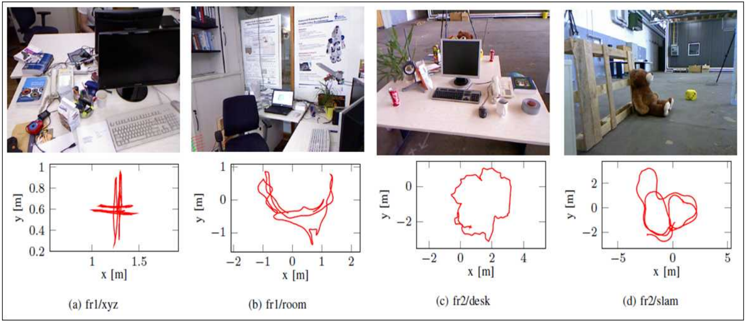
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

A Benchmark for RGB-D Tracking



Sturm, Engelhard, Endres, Burgard, Cremers, IROS 2012

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Combining Photometric and Geometric Consistency

Kerl, Sturm, Cremers, IROS 2013 propose an extension of the RGB-D camera tracker which combines color consistency and geometric consistency of subsequent RGB-D images.

Assuming that the vector $r_i = (r_{ci}, r_{zi}) \in \mathbb{R}^2$ containing the color and geometric discrepancy for pixel i follows a bivariate t-distribution, the maximum likelihood pose estimate can be computed as:

$$\min_{\xi \in \mathbb{R}^6} \sum_i w_i \, r_i^\top \Sigma^{-1} r_i,$$

with weights w_i based on the student t-distribution:

$$w_i = \frac{\nu + 1}{\nu + r_i^\top \Sigma^{-1} r_i}.$$

This nonlinear weighted least squares problem can be solved in an iteratively reweighted least squares manner by alternating a Gauss-Newton style optimization with a re-estimation of the weights w_i and the matrix Σ.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Loop Closure and Global Consistency

When tracking a camera over a longer period of time, errors tend to accumulate. While a single room may still be mapped more or less accurately, mapping a larger environment will lead to increasing distortions: Corridors and walls will no longer be straight but slightly curved.

A remedy is to introduce pose graph optimization and loop closuring, a technique popularized in laser-based SLAM systems. The key idea is to estimate the relative camera motion $\hat{\xi}_{ij}$ for any camera pair i and j in a certain neighborhood. Subsequently, one can determine a globally consistent camera trajectory $\xi = \{\xi_i\}_{i=1..T}$ by solving the nonlinear least squares problem

$$\min_{\xi} \sum_{i \sim j} \left(\hat{\xi}_{ij} - \xi_i \circ \xi_j^{-1} \right)^\top \Sigma_{ij}^{-1} \left(\hat{\xi}_{ij} - \xi_i \circ \xi_j^{-1} \right),$$

where Σ_{ij}^{-1} denotes the uncertainty of measurement $\hat{\xi}_{ij}$. This problem can be solved using, for example, a Levenberg-Marquardt algorithm.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

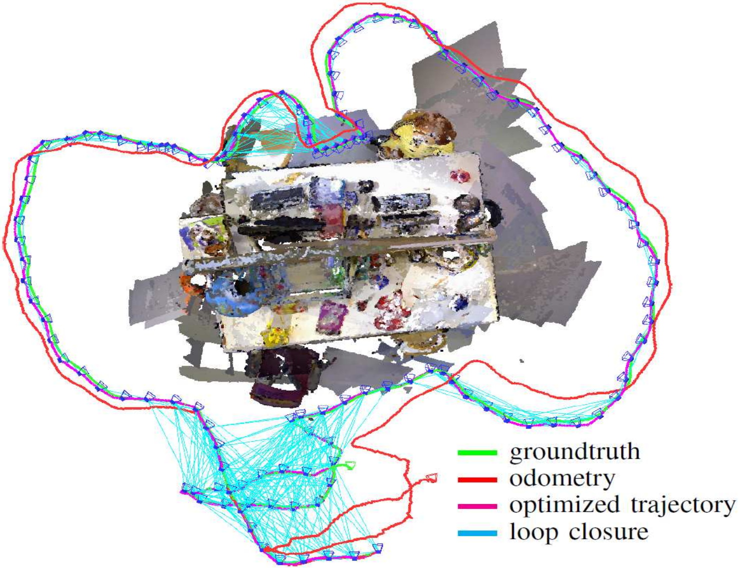
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Pose Graph Optimization and Loop Closure



Kerl, Sturm, Cremers, IROS 2013

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Dense Tracking and Mapping

Newcombe, Lovegrove & Davison (ICCV 2011) propose an algorithm which computes both the geometry of the scene and the camera motion from a direct and dense algorithm.

They compute the inverse depth $u = 1/h$ by minimizing a cost function of the form

$$\min_u \sum_{i=2}^{n} \int_\Omega \left| I_1(x) - I_i\left(\pi g_i\left(\frac{x}{u}\right)\right)\right| \, \mathrm{d}x \; + \; \lambda \int_\Omega \rho(x)\, |\nabla u| \, \mathrm{d}x,$$

for fixed camera motions g_i. The function ρ introduces an edge-dependent weighting assigning small weights in locations where the input images exhibit strong gradients:

$$\rho(x) = \exp\left(-|\nabla I_\sigma(x)|^\alpha\right).$$

The camera tracking is then performed with respect to the textured reconstruction in a manner similar to Steinbrücker et al. (2011). The method is initialized using feature point based stereo.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

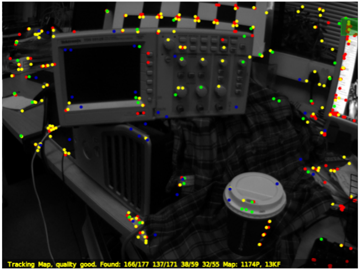
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Dense Tracking and Mapping



Newcombe, Lovegrove & Davison (ICCV 2011)

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Large-Scale Direct Monocular SLAM

A method for real-time direct monocular SLAM is proposed in Engel, Sturm, Cremers, ICCV 2013 and Engel, Schöps, Cremers, ECCV 2014. It combines several contributions which make it well-suited for robust large-scale monocular SLAM:

- Rather than tracking and putting into correspondence a sparse set of feature points, the method estimates a semi-dense depth map which associates an inverse depth with each pixel that exhibits sufficient gray value variation.
- To account for noise and uncertainty each inverse depth value is associated with an uncertainty which is propagated and updated over time like in a Kalman filter.
- Since monocular SLAM is invariably defined up to scale only, we explicitly facilitate scaling of the reconstruction by modeling the camera motion using the Lie group of 3D similarity transformations $Sim(3)$.
- Global consistency is assured by loop closuring on $Sim(3)$.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Tracking by Direct $\mathfrak{sim}(3)$ Image Alignment

Since reconstructions from a monocular camera are only defined up to scale, Engel, Schöps, Cremers, ECCV 2014 account for rescaling of the environment by representing the camera motion as an element in the Lie group of 3D similarity transformations $Sim(3)$ which is defined as:

$$Sim(3) = \left\{ \begin{pmatrix} sR & T \\ 0 & 1 \end{pmatrix} \text{ with } R \in SO(3),\ T \in \mathbb{R}^3,\ s \in \mathbb{R}_+ \right\}.$$

One can minimize a nonlinear least squares problem

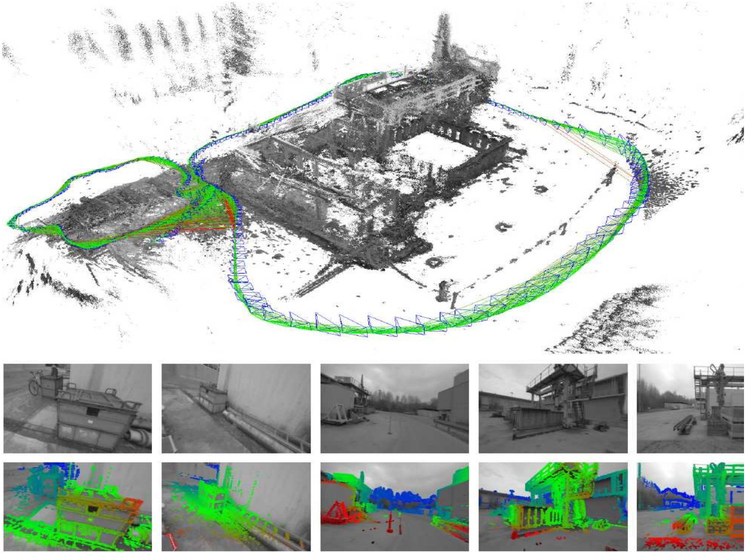
$$\min_{\xi \in \mathfrak{sim}(3)} \sum_i w_i r_i^2(\xi),$$

where r_i denotes the color residuum across different images and w_i a weighting as suggested in Kerl et al. IROS 2013.

The above cost function can then be optimized by a weighted Gauss-Newton algorithm on the Lie group $Sim(3)$:

$$\xi^{(t+1)} = \Delta_\xi \circ \xi^{(t)}, \quad \text{with } \Delta_\xi = \left(J^\top W J\right)^{-1} J^\top W\, r,\ \ J = \frac{\partial r}{\partial \xi}$$

Large-Scale Direct Monocular SLAM



Engel, Schöps, Cremers, ECCV 2014

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

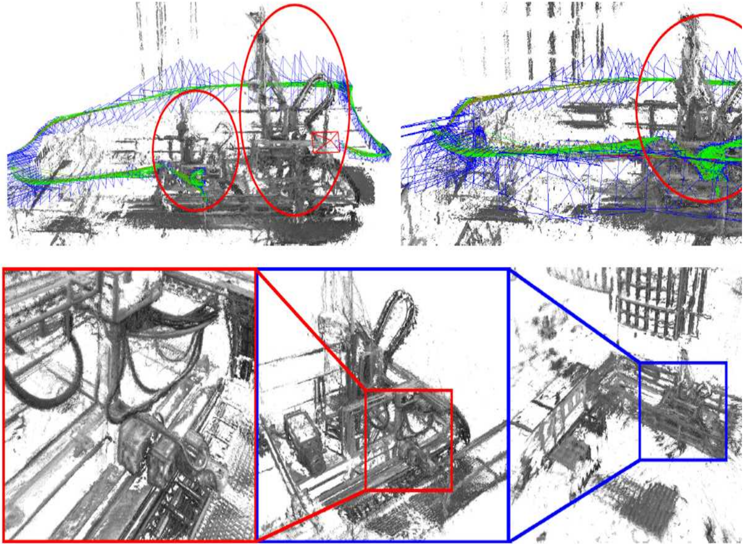
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Large-Scale Direct Monocular SLAM



Engel, Schöps, Cremers, ECCV 2014

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Towards Direct Sparse Odometry

Despite its popularity, LSD SLAM has several shortcomings:

- While the pose graph optimization allows to impose global consistency, it merely performs a joint optimization of the extrinsic parameters associated with all keyframes. In contrast to a full bundle adjustment, it does not optimize the geometry. This is hard to do in realtime, in particular for longer sequences.
- LSD SLAM actually optimizes two different cost functions for estimating geometry and camera motion.
- LSD SLAM introduces spatial regularity by a spatial filtering of the inverse depth values. This creates correlations among the geometry parameters which in turn makes Gauss-Newton optimization difficult.
- LSD SLAM is based on the assumption of brightness constancy. In real-world videos, brightness is often not preserved. Due to varying exposure time, vignette and gamma correction, the brightness can vary substantially. While feature descriptors are often invariant to these changes, the local brightness itself is not.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

From Brightness Constancy to Irradiance Constancy

log(irradiance B)   Vignette V   hardware gamma G

Brightness variations due to vignette, gamma correction and exposure time can be eliminated by a complete photometric calibration:

$$I(x) = G\Big(t\, V(x) B(x)\Big)$$

where the measured brightness I depends on the irradiance B, the vignette V, the exposure time t and the camera response function G (gamma function). G and V can be calibrated beforehand, t can be read out from the camera.

Engel, Koltun, Cremers, PAMI 2018

Windowed Joint Optimization

A complete bundle adjustment over longer sequences is difficult to carry out in realtime because the number of 3D point coordinates may grow very fast over time. Furthermore new observations are likely to predominantly affect parameters associated with neighboring structures and cameras. For a given data set, one can study the connectivity graph, i.e. a graph where each node represents an image and two nodes are connected if they look at the same 3D structure.

Direct Sparse Odometry therefore reverts to a windowed joint optimization, the idea being that from all 3D coordinates and camera frames only those in a recent time window are included. The remaining ones are marginalized out.

If one avoid spatial filtering and selects only a sparser subset of points, then the points can be assumed to be fairly independent. As a result the Hessian matrix becomes sparser and the Schur complement can be employed to make the Gauss-Newton updates more efficient.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

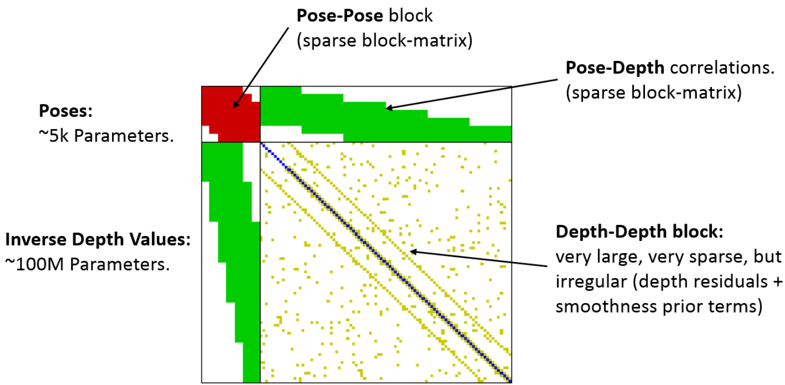
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Effects of Spatial Correlation on the Hessian



(theoretical) Hessian Structure of LSD-SLAM

=> cannot be optimized in real-time

Engel, Koltun, Cremers, PAMI 2018

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

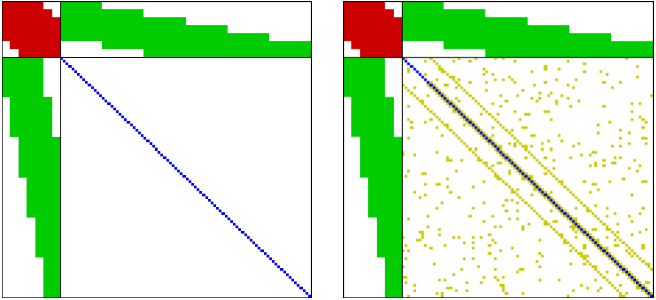
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Effect of Spatial Correlation on the Hessian Matrix



geometry not correlated | geometry correlated

Engel, Koltun, Cremers, PAMI 2018

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

The Schur Complement Trick

Solving the Newton update step (called normal equation)

$$H\,x = \begin{pmatrix} H_{\alpha\alpha} & H_{\alpha\beta} \\ H_{\alpha\beta}^\top & H_{\beta\beta} \end{pmatrix} \begin{pmatrix} x_\alpha \\ x_\beta \end{pmatrix} = \begin{pmatrix} g_\alpha \\ g_\beta \end{pmatrix},$$

for the unknowns x_α and x_β is usually done by QR decomposition for large problems. In this case, however, $H_{\beta\beta}$ is typically block diagonal (and thus easy to invert).
Left-multiplication with the matrix

$$\begin{pmatrix} I & -H_{\alpha\beta}H_{\beta\beta}^{-1} \\ 0 & I \end{pmatrix},$$

leads to:

$$\begin{pmatrix} S & 0 \\ H_{\alpha\beta}^\top & H_{\beta\beta} \end{pmatrix} \begin{pmatrix} x_\alpha \\ x_\beta \end{pmatrix} = \begin{pmatrix} g_\alpha - H_{\alpha\beta}H_{\beta\beta}^{-1}g_\beta \\ g_\beta \end{pmatrix},$$

where $S = H_{\alpha\alpha} - H_{\alpha\beta}H_{\beta\beta}^{-1}H_{\alpha\beta}^\top$ is the Schur complement of $H_{\beta\beta}$ in H. It is symmetric, positive definite and block structured. The equation $Sx_\alpha = \ldots$ is the reduced camera system.

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

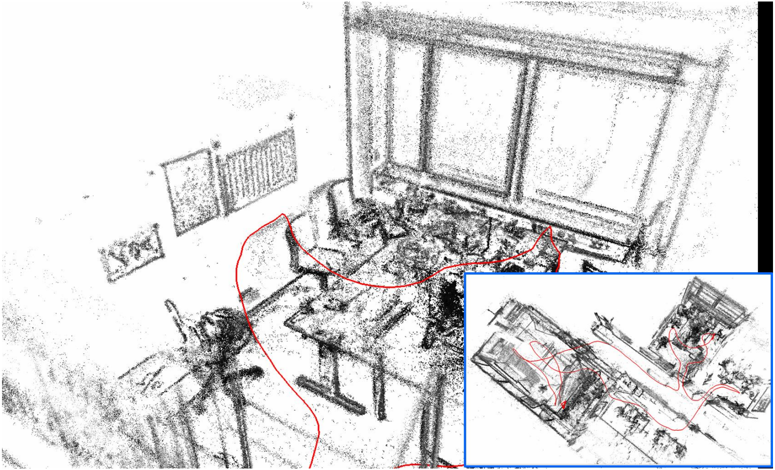
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Direct Sparse Odometry



Engel, Koltun, Cremers, PAMI 2018

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

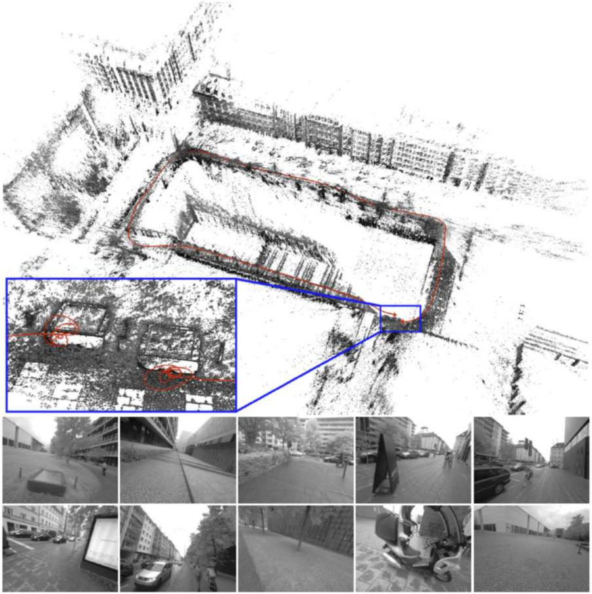
Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Direct Sparse Odometry



Engel, Koltun, Cremers, PAMI 2018

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

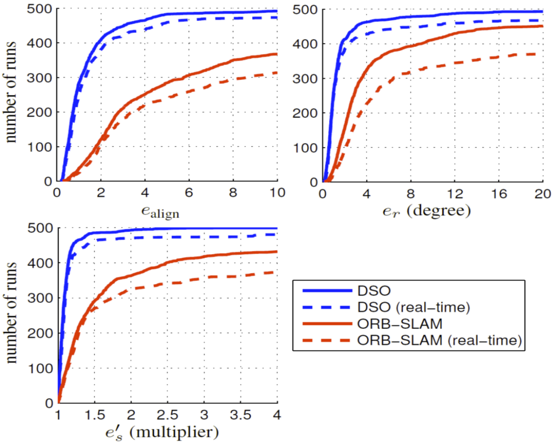
Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry

Quantitative Evaluation

A quantitative comparison of Direct Sparse Odometry to the state-of-the-art keypoint based technique ORB SLAM shows substantial improvements in precision and robustness:



# of runs with a given error in translation, rotation and scale drift.

Engel, Koltun, Cremers, PAMI 2018

Direct Methods

Realtime Dense Geometry

Dense RGB-D Tracking

Loop Closure and Global Consistency

Dense Tracking and Mapping

Large Scale Direct Monocular SLAM

Direct Sparse Odometry