2018 September Reading Reports

CodeSLAM

It introduced a new compact but dense representation of scene geometry which is conditioned on the intensity data from a single image and generated from a code. For mapping, photometric error and geometric error are related not only the pixel position and the pose of the camera, but also the **codes**, which will be optimized. For tracking or localization, it uses geometric cost only and with coarse to fine optimization. The SLAM system is inspired by PTAM. The overall results are not as good as VI method, but performs respectably for a vision only system.

Learning Compact Geometric Feature

Local features in point cloud is learned in this work.

First, it parameterizes the points within the sphere by spherical histogram (binned in $N = R \times E \times A$), which results in the feature $\mathbb{R}^N$.

Triple embedding loss is used to train the model mapping the high-dimensional spherical histogram's space to a very low-dimensional Euclidean space. The training triples are generated by two concentric spheres, which are used to determine the positive and difficult negative samples of points (in training, negative can also be from an entire model). Synthesizing depth images are used to automatically generate enough training data.

Correspondences are found by kd-tree. It can serve as drop-in replacements for existing descriptors.

The transformation is given in some of the evaluations.

Pipeline of **fast global registration** (FGR, original with FPFH) is adapted with the proposed method to register the points. Robust optimization can prune false positives.

Fast Global Registration

It **does not require initialization** and can align noisy partially overlapping surfaces at a **computational cost** that is more than an order of magnitude lower than the global alignment + refinement framework. It can also jointly align multiple partially overlapping surfaces directly by a single optimization of a global objective.

Related Work

Typically workflow consists of two stages: global alignment, followed by local refinement. The global alignment includes the methods operate on candidate correspondences (RANSAC, pose clustering, etc., samples), the branch-and-bound framework (Go-ICP). The local refinement is usually done by ICP and its variants.

Objective

A scaled Geman-McClure estimator:

$$\rho(x) = \frac{\mu x^2}{\mu + x^2}$$

can be solved by Black-Rangarajan duality. It can be optimized by alternating between $\mathbf{T}$ and $\mathbb{L}$ (iteration needed, but very efficient). Reciprocity test and tuple test are used to improve the inlier ratio. Multi-way registration is similar with the original version, added backbone odometry terms to stabilize the optimization. Correspondences are not changed in the optimization.

Robust Reconstruction of Indoor Scenes

It presents a dedicated formulation for dense surface reconstruction that identifies outliers by directly optimizing for surface alignment, using an objective that efficiently incorporates dense correspondence constraints.

Fragment is constructed by visual odometry, geometric registration for each pair is run for find loops, robust optimization is used to handle with the outliers, and final model is refined using ICP and volumetric integration.

Geometric registration determine the loops, and robust optimization with **line process** is used to solve the objective function. Proximity of correspondence pairs is used to simplify the objective. And finally becomes a quadratic form, which can be solved by **g2o**. BRE (???) is used to evaluate the results (Amazon Mechanical Turk).

Stereo Vision-based Semantic 3D Object and Ego-motion Tracking for Autonomous Driving

Ego-motion and car motion are estimated in the work. It mainly uses non-linear optimization (BA) for both of the estimation. Learning-based semantic labelling (car detection), 2D-3D car representation (classification in 2*8 = 16 viewpoints), feature tracking (ORB for temporal matching, the center distance and shape similarity of the 2D boxes) and car kinematics model are used. Further alignment with stereo point-cloud and the estimated points is carried out.

LIPS: LiDAR-Inertial 3D Plane SLAM

Closest Point Plane Representation

It is based on plane representation on $\mathbf{n}, d$. The optimization includes the plane representations and the states.

In real tests, RANSAC plane extraction is needed, the small-scale tests show the proof-of-concept only.

Self-Supervised Sparse-to-Dense: Self-Supervised Depth Completion from LiDAR and Monocular Camera

Three kinds of loss are designed to make sparse point-cloud into dense point-cloud, depth loss, photometric loss and smooth loss.