2018 March Reading Reports

iBoW-LCD

Incremental Bags of Binary Words with inverted index and concept of islands etc. The ideas in LCD part can be adopted in further application.

Different from previous offline bag of words, this method using online incremental BoW for loop closure detection. The words in the paper is binary, thus an algorithm based on Muja's hierarchical structure of indexing and matching binary features is adopted and be modified into incremental fashion. This structure has the ability to add or delete the words online. The tests in the paper use ORB descriptors.

Incremental Segment-Based Localization in 3D Point Clouds

The most different parts of this paper and Segmatch are the incremental methods for the dynamic voxel grid method, the calculation of the normals, the curvatures, the region growing segmentation, and a faster geometric validation method based on maximum pairwise consistent set (MPCS).

NICP

First, it introduces projection and unprojection of a 3D point cloud onto a range image.

The surface normal is calculated by integral images. And for reducing the noise of plane, the ellipsoid is changed as the same as what is mentioned in GICP.

Then points are projected into index image. For the current points, it only computed only; while for the reference points, the index image is calculated with actual estimate T. Then, some correspondences are found as candidates, and some candidates are discarded by distance, curvature and normal angle.

To determine the transformation, it uses 6 length vectors and its corresponding information matrix, the local parameterization of the perturbation is used to do calculation iteratively.

KineticFusion

Dense Surface

Four parts, Surface measurement, surface reconstruction update, surface prediction, sensor pose estimation.

For the surface measurement, noise of points are reduced in depth map and back-project the filtered one into the sensor's frame. The normal of each pixel is calculated by two of the neighbor points. Then multi-scale pyramid depth map (vertex) and normal map are calculated.

Volumetric truncated signed distance function (TSDF) is used to present the plane in the method, for the points above the surface μ is used to truncate the value; the points behind μ will be null. Raw depth data and weights are applied for fusing multiple measurements.

$\mathbf{p}$ is orthogonal to the zero level set, the normal on the plane is calculated by numerical derivative of the SDF. The range of the sensor is limited, thus it keeps the calculation as constant. A simple approximation is used for the higher quality intersections.

For the final pose estimation, model prediction is used instead of just one frame. Fast projective data association algorithm to obtain correspondences. Some outliers are filtered based on some features, e.g. the mask, the distance between model and scan, and the normal orientation difference. Then a ICP like method based on the point-plane metric is used for the final results, with the help of pyramids from coarse to fine framework.

Finally, a stability and validity check, i.e null space of the normal system, and magnitude of incremental transform parameters makes only good tracking to update the model.

DVO

The name is from NICP as dense visual odometry. Assuming a static scene, and the twist is constant, and the temporal derivative of the image is constant.

Energy minimization or maximize photoconsistency on the image warps, which is like the unprojection into the image frame (only the 3D points part of the image). The final error function can be consider as intensity differences between two images, and results in a least square problem with 1x6 constraints and a 6x6 normal equation. A coarse to fine refinement is applied also.

NDT

It is designed for the matching in mine tunnel.

First, subdivide the space occupied by the model into regularly sized cells (squares in 2D or cubes in 3D).

Then for the cell containing more than some minimum number of points, the mean vector and covariance matrix are calculated. And use $N(\mathbf{q}, \mathbf{C})$ to model the normal distribution. An error function is built based on the normal distribution with transformation $\mathbf{p}$. Then minimizing the error function by Newton's algorithm.

For the 3D case, the main different is the Hessian and gradient for the transformation. It is derived in the paper. Then some practical implementations of 3D-NDT are discussed, e.g. the sampling method, the cell size and the discretization methods.

ICP (iterative closest point) algorithm (Besl&McKay, 1992; Chen&Medioni, 1992)

Integral Images

Surface normal from organized point cloud data using integral images.

It is suitable for organized points. Smoothing upon different depth values, depth changes are utilized for the final smoothing area map. Two methods are proposed using the smoothed integral images. One is the smoothed depth changes (by cross product on two vectors based on the adjacent smoothed integral images) and the other is based on optimization (covariance matrices).

This method is fast and robust for organized point cloud, but it could smooths over edges.

ORB-SLAM

Orb-slam contains four major parts: **tracking, mapping, relocalization, and loop closing**. It is viewpoint and illumination invariant, and uses covisibility graph to make the tracking and mapping only related to the local covisible area. Essential graph is build from a spanning tree. Relocalization and initialization are discussed. Survival of the fittest approach is used to map point and keyframe selection, which improves tracking robustness and discards redundant keyframes. LM algorithm in g2o is used to carry out all optimizations. For **each map point** p_i, it stores its 3-D position $X_{w,i}$ in the world coordinate system, the viewing direction, a representive ORB descriptor and maximum, minimum distances at which the point can be observed. For **each keyframe** K_i is stores the camera pose T_{iw}, from the world to the camera coordinate system, the camera intrinsics and all the ORB features extracted in the frame (associated or not with a map point, undistorted). The covisibility graph is a graph whose edges connect every two keyframes sharing observations of the same map points (at least 15). An essential graph is a strong network of cameras containing subset of edges from the covisibility graph with high covisibility ($\theta_{min} = 100$), and the loop closure edges.

Initialization

Two methods are used in parallel. Extract ORB features with the finest scale. The one is the homography for the planar or low parallax cases; the other one is the fundamental matrix for the nonplanar or enough parallax cases. Finally a full BA is used to refine the initial reconstruction.

Tracking

For every frame from the camera, motion-only BA is applied. First ORB extraction using different parameters according to different pixels in the datasets. Then initial pose is estimated from previous frame using a constant velocity motion model (guided search). If this model is violated (not enough matches), a searching around map points positions in the last frame is applied (wider search). If the tracking is lost, global relocalization will work. PnP is used to find a camera pose (RANSAC). Tracking local map requires a local map containing the keyframes K_1 with the same map points in current frame ($K_{ref} \in K_1$, which share the most map points) and K_2 the neighbors with K_1. Discard the points lay out of the bounds; discard the points with too large distance between viewing directions and viewing rays; discard the points out of the scale invariance region; compute the scale and match unmatched features with the predicted scale. Keyframes: 20 frames (global relocalization and local mapping is idle or after the last keyframe insertion), 50 points, 90% points.

Local Mapping

For every new keyframe K_i. Insert new keyframe: update the covisibility graph, add a new node and edges, update the spanning tree, and the bags of words representation. Local points culling, trackable points are retained (25% visible, at least three keyframes). New map point creation, triangulating ORB from connected keyframes. Projected in the rest of connected keyframes.

Local bundle adjustment (local BA), currently processed keyframe, all the keyframes connected to it in the covisibility graph and all the map points seen by those keyframes. Local keyframe culling, detect redundant keyframes and delete them (90% in at least other three keyframes).

Loop Cloosing

The last keyframe K_i processed by the local mapping, and tries to detect and close loops. A normalizing score is used (discard the other keyframes with score lower than the lowest score of covisibility graph). It needs consecutively three loop candidates (keyframes). Similarity transformation, geometrical validation, 3-D-to-3-D correspondences for each loop candidate, RANSAC to find a similarity transformation. To fuse duplicated map points and insert new edges in the covisibility graph that will attach the loop closure. Pose graph optimization over the Essential Graph. Each map point is transformed according to the correction of one of the keyframes that observes it.

Visual-Inertial Monocular SLAM With Map Reuse

Close loops in real-time and localize the sensor reusing the map in already mapped areas. (Others like odometry, marginalize past states or need a map, which is built offline)

A reliable **visual-inertial initialization** — provides fixed states for optimization.

Zero-drift localization in the same workspace.

Expect to achievebetter accuracy and robustness by using stereo or RGB-D cameras.

Weakness of the proposed IMU initialization is that it relies on the initialization of the monocular SLAM.

Slow initialization comparing toVINS system (investigate the use of the gyroscope).

Some Ideas

LOAM algorithm extracts lines (edges) and planes from single scan. Since single scan can provide such features within local information. At the same time, the provided line or plane features can help to calculate the error of lines or planes in the following scans.

But these features may not be good for jobs related pure localization, since the features are just points with some limited information, like curvature or planarity. It is the problem about this method.

```
@inproceedings{zhang2014loam,
  title={LOAM: Lidar Odometry and Mapping in Real-time.},
  author={Zhang, Ji and Singh, Sanjiv},
  booktitle={Robotics: Science and Systems},
  volume={2},
  year={2014}
}
```

TODO

If one wants to do place recognition, one thing is to find some common features that could appear in multiple places.

```
@article{dube2016segmatch,
  title={SegMatch: Segment based loop-closure for 3D point clouds},
  author={Dub{\'e}, Renaud and Dugas, Daniel and Stumm, Elena and Nieto,
Juan and Siegwart, Roland and Cadena, Cesar},
  journal={arXiv preprint arXiv:1609.07720},
  year={2016}
}
```