2018 February Reading Reports

FAB-MAP

The mean idea of FAB-MAP is to use a probabilistic framework for appearance-based place recognition. The original paper used 128D SURF descriptors extracted from images. And then cluster them to 11k words as vocabulary. Then compute Chow Liu tree to approximate the probability of $p(Z_k|L_i, Z^{k-1})$ instead of a Naive Bayesian assumption.

Hidden events e_i are introduced to provide ability to incorporate multiple sensors and factor the distribution $p(Z|L_j)$ into two parts.

It deals with the case that a new place could appear, and the corresponding probability under such case. As mentioned by the authors, Monte Carlo approximation, or a sampling method, is applied to get a better performance.

Location Prior should be given.

Smoothing is applied for rejecting a single similar image pair.

The assumptions in this methods are:

1. Sets of observations are conditionally independent given position: $p(Z_k|L_i, Z_{k-1}) = p(Z_k|L_i)$.
2. Detector behavior is independent of position: $p(z_j|e_j, L_i) = p(z_j|e_j)$.
3. Location models are generated independently by the environment.
4. Approximation of $p(e_j|Z_k) \approx p(e_j|z_{j_k})$.

Some comments:

1. The method needs off-line training, which requires some conditions, e.g. no overlapping on the training data.
2. The words are fixed in the origin work, which are obtained from descriptors in training data.
3. Some hyper-parameters, like priors could be hard to determine.

```
@article{cummins2008fab,
  title={FAB-MAP: Probabilistic localization and mapping in the space of
appearance},
  author={Cummins, Mark and Newman, Paul},
  journal={The International Journal of Robotics Research},
  volume={27},
  number={6},
  pages={647--665},
  year={2008},
  publisher={SAGE Publications Sage UK: London, England}
}
```

PL-SLAM

Visual SLAM with point and line features. The paper used minimum four parameters to describe line features and proposed some method like merge lines, calculate Jacobian matrix for fast computation, and modified ORB-SLAM2 to fit with the proposed method.

```
@article{zuo2017robust,
  title={Robust Visual SLAM with Point and Line Features},
  author={Zuo, Xingxing and Xie, Xiaojia and Liu, Yong and Huang,
Guoquan},
  journal={arXiv preprint arXiv:1711.08654},
  year={2017}
}
```

LocNet

First, extract some features based on ring data, which is possible to be rotation invariant , and kind of translation invariant (from consecutive points). Then a siamese network is used to compare the distance of two frames. Contrastive loss function is proposed to train the data here. The prior global map is utilized by a Kd-tree for fast matching. The result from matching is fused by odometry information, also a resamping for rotation estimation (wrong observation with odometry information will decrease the probability). Finally, the previous results are use as initial value for ICP algorithm.

In trainning, use tw consecutive frames as the same place. Distance greater than p will be taken as negative samples.

The problem of the method can be as following:

1. It cannot work on 3D environment with height changes.
2. Can only work on a place with map, i.e. localization work only. If the robots move to some places, it will not update the map.

3. Some parameter related performance.
4. The comparision of segmatch is based on the same sequence, which is a point to be attacked easily.
5. Odometry is necessary in this method currently.

```
@article{yin2017locnet,
  title={LocNet: Global localization in 3D point clouds for mobile
robots},
  author={Yin, Huan and Wang, Yue and Tang, Li and Ding, Xiaqing and
Xiong, Rong},
  journal={arXiv preprint arXiv:1712.02165},
  year={2017}
}
```

BOW

The main idea of bag of words methods is extract some features from images. Then cluster them into several clusters as words. In the cited paper, it first use kmeans++ to get the words. A vocabulary tree is used to fast search. TF-IDF then is applied for calculating the weights of the words in a image. The final vectors to be compared consist of the weights of the words. It uses inverse and direct index, which helps to find similar images and find the corresponding features faster.

```
@article{galvez2012bags,
  title={Bags of binary words for fast place recognition in image
sequences},
  author={G{\'a}lvez-L{\'o}pez, Dorian and Tardos, Juan D},
  journal={IEEE Transactions on Robotics},
  volume={28},
  number={5},
  pages={1188--1197},
  year={2012},
  publisher={IEEE}
}
```

IMLS-SLAM

A new scan-to-model framework using an implicit (**Implicit Moving Least Squares (IMLS) surface representation**) surface representation of the map.

First, we compute a local de-skewed point cloud from one rotation of the 3D LiDAR. Second, we select specific samples from that point cloud to minimize the distance to the model cloud in the third part.

Part 1, the de-skewed point cloud is from previous odometry and registered into the end of the point cloud. Then voxel growing is applied for ground points removal. Then small clusters are removed and ground is added back.

Part 2, choose $9s$ samples from the results from results of part 1. The choosing of samples is based on point features, which are aimed at different translation or rotation parameters.

Part 3,

A scan-to-model matching using implicit moving least square representation. Calculate the distance between implicit surface and points by $I^{P_k}(x)$. Since mininizing $\sum_{x_j \in \tilde{S}_k} |I^{P_k}(x)(Rx_j + t)|^2$ cannot be approximate by a linear least-square problem. A projection of points on IMLS is calculated, and the points are $Y_k : y_j = x_j - I^{P_k}(x)\overrightarrow{n_j}$. Then minimize error $\sum_{x_j \in \tilde{S}_k} |\overrightarrow{n_j}(Rx_j + t - y_j)|^2$, which stands for the differences between x_j and y_j. After one iteration, new projection and transformation is calculated.

Comments:

1. Not real-time.
2. Some procedures in the pipeline seems not very useful, like the removal of dynamic objects
3. Many we can test on the mapping job for JD, which is based on the merged local map.