

NAVER LABS Mapping & Localization Challenge Indoor - 2nd place

Hyunyoung Jung

gusdud1500@snu.ac.kr

CMALAB, Seoul National University

1. Overview

정확한 Localization을 구현하기 위하여 전체적인 과정은 [1]의 파이프라인을 따르고 있다. 테스트 시 쿼리 이미지가 주어졌을 때 해당 이미지에 대하여 Global descriptor를 계산한 후, 데이터베이스 이미지들을 검색하여 쿼리 이미지와 가장 유사한 N개의 이미지를 고른다.(Section. 2) 이후에 Local descriptor를 사용하여 이전 과정에서 선택된 데이터베이스 이미지와 쿼리 이미지 간의 correspondence points를 검출 후 geometric verification 단계를 통해 부정확한 매칭점들을 제거한다.(Section. 3) 이후 데이터베이스 이미지와 인접한 타임스탬프를 갖는 Lidar points들을 projection한 후 해당 매칭점들과 가장 가까운 거리에 존재하는 lidar point를 해당 key point에 대한 3D 포인트로 설정을 한다. 이렇게 선택된 3D Point들과 매칭된 쿼리 이미지 포인트 간에 perspective-n-point 알고리즘을 통해 최종 pose를 계산한다.(Section. 4)

2. Image Retrieval

먼저 데이터베이스 이미지들에 주어진 camera pose(translation)정보를 기반으로 K-nearest neighbor 알고리즘을 적용하여 2개의 cluster로 분류하였다. 이후 기 학습된 모델 [2, 3]을 통해 데이터베이스 및 쿼리 이미지에 대한 Global descriptor를 계산한 후 각각의 클러스터에서 30장의 가장 유사한 이미지를 선택한다.

3. Feature matching & Reranking

앞서 선택된 총 60장(/ 쿼리 이미지)의 이미지에서 [4, 5] 모델을 이용해 key point와 매칭점들을 계산하고 선택된 포인트들 간의 Ransac을 이용한 Fundamental matrix estimation을 통해 outlier를 제거한다. 이렇게 최종적으로 남은 포인트들의 개수(inlier)을 이용하여 60장의 데이터베이스 이미지를 정렬하여 가장 inlier수가 많은 순서대로 하나의

쿼리 이미지에 대해 총 20장의 데이터베이스 이미지를 선택한다. Inlier의 개수가 같을 경우에는 global descriptor로 계산한 similarity의 순서를 따른다.

4. 3d Projection & pose estimation

이전 단계에서 선택된 20장의 데이터베이스 이미지에 대해 해당 이미지에 인접한 timestamp를 갖는 Lidar point들을 projection한다(-4 ~ +4 seconds). 이후 projected된 포인트들과 데이터베이스 이미지의 선택된 키포인트 간의 거리가 가장 최소가 되는 점들을 찾아 해당 lidar point들을 해당 데이터베이스 포인트의 3D Point로 가정한다. 이 때, projected lidar point와 데이터베이스 key point간의 거리가 7.5 pixel 이상 차이가 날 경우에는 제외한다.

이렇게 선택된 3D point들을 Section. 3에서 매칭된 쿼리 이미지 상의 키 포인트와 함께 Ransac을 이용한 PnP 알고리즘을 통해 쿼리 이미지에 대한 최종 pose를 계산한다.

Reference

[1] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, Marcin Dymczyk . From Coarse to Fine: Robust Hierarchical Localization at Large Scale

[2] Albert Gordo, Jon Almazan, Jerome Revaud, Diane Larlus, End-to-end Learning of Deep Visual Representations for Image Retrieval, IJCV 2017

[3] Jerome Revaud, Jon Almazan, Rafael S. Rezende, Cesar de Souza , Learning with Average Precision: Training Image Retrieval with a Listwise Loss, ICCV 2019

[4] Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich, SuperPoint: Self-Supervised Interest Point Detection and Description.

[5] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich, SuperGlue: Learning Feature Matching with Graph Neural Networks