

NAVER LABS Mapping & Localization Challenge

세종대학교 RCV 연구실, 김지원, 김태주, 황유진
{jwkim, tjkim, yjhwang}@rcv.sejong.ac.kr

본 챌린지의 아웃도어 데이터셋 대회는 차량에 부착된 RGB-Stereo 카메라를 이용하여 차량의 현재 위치(6DoF)를 인식하는 것을 목적으로 한다. 한번 이상 주행을 마친 지도 영상(이하 DB)에는 각 영상의 절대좌표정보가 함께 기록되어 있어, DB 영상과 현재 위치 영상(이하 QR) 간 상대적인 위치 계산을 통해 QR의 절대좌표를 알 수 있다. 본 대회의 문제 해결을 위해 사용한 SejongRCV-Outdoor 팀의 파이프라인은 그림1과 같다. 팀의 알고리즘은 크게 3개의 모듈로 구성되어 있으며, 첫번째 모듈은 QR 이미지와 가장 유사한 장소를 지도 DB 내 에서 찾아내는 'Place Recognition' 모듈, QR과 DB간 2D-3D 대응점 관계를 Perspective-n-Point (PnP)로 풀어 QR의 위치와 방향 정보(6DoF)를 추정하는 'Pose Estimation' 모듈, 마지막으로 위치인식 정확도 향상을 위한 'Visual odometry' 모듈이 해당된다.

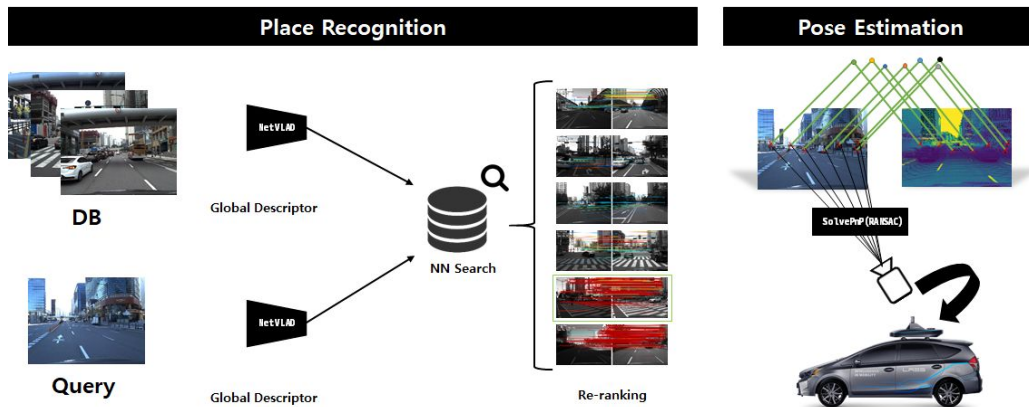


그림1. SejongRCV outdoor 팀의 문제 해결 전체 흐름도

1. Retrieval based Place Recognition

Image Retrieval기반의 Place Recognition 문제를 해결하기 위하여 영상 기술자인 Global descriptor로써 64개의 cluster로 구성된 NetVLAD[1]를 사용하였다. NetVLAD 구현에는 GitHub에 공개된 pytorch-NetVlad¹ 코드를 베이스로 하였고, Repository에 공개된 pre-trained weight (NetVLAD:Pittsburgh 250k)를 추가 학습 없이 사용하였다. NetVLAD를 통해 DB 이미지의 Global descriptor 를 추출 및 저장하고, QR 이미지와 유사한 DB 영상 검색(searching)을 위해 sklearn의 Nearest Neighbors 함수를 이용하였다.



그림 2. QR에 대한 Place recognit Top-10 결과 (Green-Box: Positive , Red-Box: Negative)

DB로 부터 QR와 가장 유사한 영상 10개를 선별하기 위하여 Re-ranking 기법을 사용하였으며, Re-ranking과정에는 올해 6월 개최된 International Conference on Computer Vision and Pattern Recognition(CVPR) 에서 발표된 SuperGlue[2] 방법론을 활용하였다. SuperGlue는 딥러닝 기반 local

¹ <https://github.com/Nanne/pytorch-NetVlad>

feature 알고리즘인 SuperPoint[3]를 통해 키포인트를 추출하고, Graph Neural Network을 이용해 키포인트를 매칭하는 알고리즘이다. SejongRCV-Outdoor 팀은 QR 영상과 후보 이미지 간 키포인트 매칭 갯수를 기준으로 Re-ranking 수행하였으며, 최종적으로 DB 영상 하나를 정답으로 선정하였다. 이 때, 실험에는 SuperGlue 저자가 공개한 SuperGlue Pretrained Network(Outdoor training) 코드²를 사용하였다.



그림 3. Re-ranking 과정을 통해서 선별한 QR(좌)-DB(우) Pair 예시

2. Pose Estimation

첫번째 모듈을 통해 선정한 Top-1 DB 영상과 해당 DB의 3D 포인트를 이용해 월드좌표계를 기준으로 QR에서의 차량의 위치(x, y, z)와 쿼터니언 값(q_w, q_x, q_y, q_z)를 추정하였다. DB 영상 마다 제공되는 3D 포인트와 차량의 위치정보(6DoF) 파싱은 네이버랩스에서 제공하는 튜토리얼 코드³를 참고하였으며, 해당 코드에서는 multiple Lidar 프레임을 쌓는 방법과 3D 포인트를 2D 이미지에 투영(projection)시키는 방법도 함께 제공하고 있다. 튜토리얼코드에서는 3차원 라이다 정보의 신뢰 구간을 5m 이상으로 설정하였으나, 본 팀은 5m~150m 로 설정하여 해당 구간 내의 3차원 포인트를 정밀 자세 추정에 사용하였다.

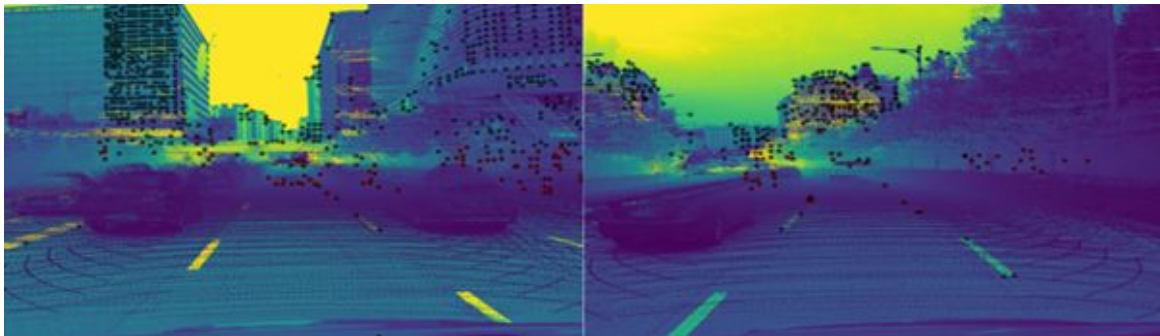


그림 4. 투영된 Lidar 포인트와 매칭 포인트(검정) 자세 추정에 사용된 포인트(빨강)

Pose Estimation의 솔루션으로 Perspective-n-Point(PnP) 알고리즘을 사용하였으며, 구현체로는 opencv에서 제공하는 solvePnP Ransac 를 사용하였다. 해당 방법론은 RANSAC으로 노이즈를 제거하고 영상의 2D 포인트와 그에 대응하는 3D 포인트의 관계를 이용하여 카메라의 pose를 알아내는 알고리즘이며, 환경 강인성 확보를 위해 solvePnP Ransac 함수의 알고리즘 선택 옵션 중 AP3P[4] 방법론을 사용하였다.

자세 추정 알고리즘을 구체적으로 설명하면 다음과 같다. SuperGlue 알고리즘으로 QR 영상과 Top-1 DB 영상의 매칭 포인트(x, y)를 추출한다. 그리고 3D 포인트를 DB에 투영(projection)시킨 후 2D 영에서 매칭된 키포인트와 투영된 3D 포인트가 1픽셀 이내인 경우에만 해당 3D 포인트를 2D 포인트와 대응관계에 있다고 정의하였다. 이러한 2D-3D 대응점을 PnP 알고리즘에 적용하여 (QR 영상에서의) 카메라 pose를 추정하고 여기에 차량의 중심 좌표계와 카메라 좌표계 사이의 변환 행렬을 곱하여 최종적인 차량의 위치(x, y, z)와 쿼터니언 값(q_w, q_x, q_y, q_z)를 추정하였다.

² <https://github.com/magicLeap/SuperGluePretrainedNetwork>

³ https://github.com/naverlabs/mapping-and-localization-challenge/blob/master/labs_outdoor_dataset_tutorial.ipynb

3. Visual Odometry

Place Recognition 과정에서 QR 이미지와 가장 유사한 DB 영상을 성공적으로 찾아냈을 때, Pose Estimation의 결과값을 신뢰할 수 있다. 하지만 QR 영상과 유사한 DB영상이 반드시 존재하는 것을 보장할 수 없고, 실제 시각적으로 유사한 이미지라 할지라도 매칭된 키포인트의 수가 매우 적은 경우도 존재한다. 이러한 문제는 최종 Pose Estimation의 성능저하를 야기하며, 본 팀은 이 문제를 해결하기 위해 제공되는 50장의 시퀀스 이미지 모두를 사용하여 Place Recognition을 수행하였고, 가장 많은 키포인트가 매칭된 QR-DB pair를 선정하였다. 해당 QR-DB pair에서의 QR의 차량 자세를 추정한 이후 Visual Odometry를 사용해 마지막 QR에 대한 차량의 자세를 추정하였다. 이러한 방법으로 마지막 QR와 유사한 DB를 찾을 수 없는 상황에서도 앞선 QR 영상에서 유사한 DB를 찾을 수 있었고, 해당 지점부터 Visual Odometry를 통해 정확한 값을 예측할 수 있었다. 본 챌린지를 위해 사용된 코드와 각종 설정 파라미터는 SejongRCV GitHub⁴에 공개하였다.

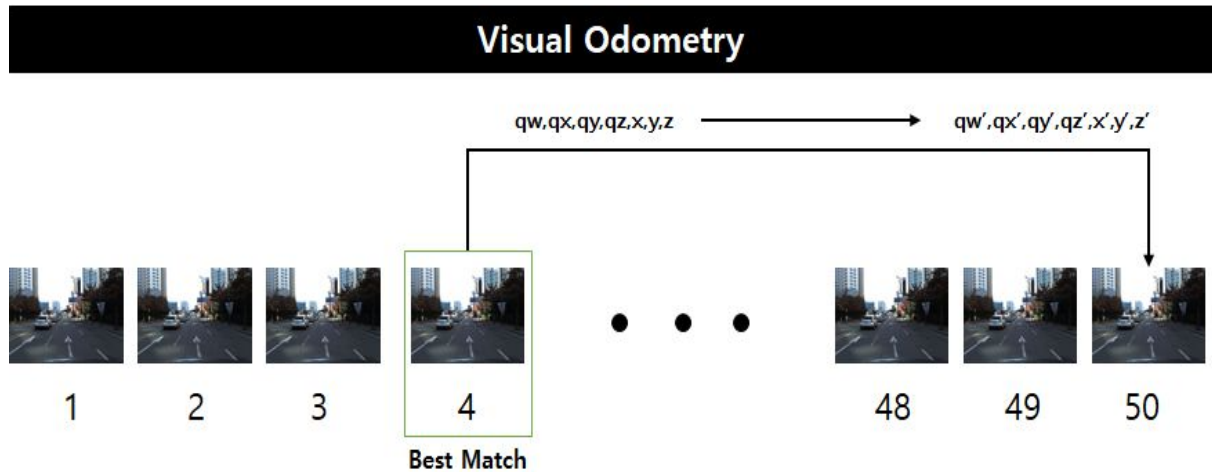


그림 5. 50개 모든 QR에 대해서 Place Recognition 수행 후 Best match를 선정하여 해당 이미지부터는 Visual Odometry를 통해 최종 50번째 QR를 추정하였다.

4. Reference

- [1] Arandjelovic, Relja, et al. "NetVLAD: CNN architecture for weakly supervised place recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [2] Sarlin, Paul-Edouard, et al. "Superglue: Learning feature matching with graph neural networks." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
- [3] DeTone, Daniel, Tomasz Malisiewicz, and Andrew Rabinovich. "Superpoint: Self-supervised interest point detection and description." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018.
- [4] Ke, Tong, and Stergios I. Roumeliotis. "An efficient algebraic solution to the perspective-three-point problem." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

⁴ <https://github.com/sejong-rcv/SejongRCV-Outdoor>