

# 단일 RGB 영상을 이용한 비주얼 오도메트리

이주성, 황상원, 김우진, 이상윤\*

e-mail: {m3155, sangwon1042, woojinkim0207, syleee}@yonsei.ac.kr\*

## RGB-VO: Visual Odometry using mono RGB

Joosung Lee, Sangwon Hwang, Woo Jin Kim, Sangyoun Lee\*  
School of Electrical and Electronic Engineering, Yonsei University

### 요 약

주요 주행과 로봇 시스템의 기술이 발전하면서 이와 관련된 영상 알고리즘들의 연구가 활발히 진행되고 있다. 제안 네트워크는 단일 영상을 이용하여 비주얼 오도메트리를 예측하는 시스템이다. 딥러닝 네트워크로 KITTI 데이터 세트를 이용하여 학습과 평가를 하며 네트워크의 입력으로는 연속된 두 개의 프레임이 들어가고 출력으로는 두 프레임간 카메라의 회전과 이동 정보가 된다. 이를 통하여 대표적으로 자동차의 주행 경로를 알 수 있으며 여러 로봇 시스템 등에서 활용할 수 있다.

### 1. 서론

비주얼 오도메트리는 카메라를 통하여 주행 경로를 예측하는 연구이다. 최근 자율 주행 및 로봇 시스템의 연구가 활발히 진행되면서 이와 관련된 비주얼 오도메트리 기술도 점점 활발히 진행되고 있다. 기존 오도메트리 연구는 SLAM 시스템의 한 부분으로 mapping과 localization을 하는데 사용이 된다. 따라서 SLAM 시스템의 정확도는 오도메트리와 밀접한 관련이 있기에 오도메트리의 정확도는 중요하다. 또한 SLAM 시스템은 실시간 결과를 글로벌 옵티마이제이션을 통하여 계속해서 업데이트를 하는 작업을 거친다. 즉 이는 전체적인 데이터를 이용하여 과거 주행 경로를 계속해서 수정을 하게 된다. 본 논문에서는 옵티마이제이션 없이 RGB 단일 영상들을 이용하여 frame-to-frame 딥러닝 네트워크를 학습하여 실시간 비주얼 오도메트리를 예측한다.

### 2. 본론

기존의 비주얼 오도메트리는 기하학적 방법을 이용하는 VISO2-M [1]이 있다. VISO2-M은 단일 이미지 프레임에서 handcraft 특징들을 도출하여 프레임간의 회전 행렬(rotation matrix)과 병진 벡터(translation vector)를 찾아내게 된다. 하지만 이러한 기하학적 방법은 단일 영상을 사용할 시 깊이 정보가 없기 때문에 스케일에 대한 문제가 심각히 일어나게 된다. 또한 특징 매칭 시 outlier 혹은 occlusion이 일어난 경우 잘 되지 않는다. 또한 여러 파라미터를 환경에 따라 튜닝을 다르게 튜닝을 해주어야 하는 문제점이 있다.

이러한 기하학적 방법의 문제로 학습 방법이 연구되었고 학습 모델은 다양한 노이즈에 강인한 특성들을 지니게 된

다. 성공적으로 알려진 첫 번째 학습 방법은 SVR-VO [2] 네트워크로 SVM 알고리즘 [4]을 사용하여 학습한 회귀 모델이다. SVR-VO는 옵티컬 플로우를 양자화를 하여 특징을 뽑아내고 이와 카메라의 예고 모션과 학습을 진행하게 된다.

#### 2.1 제안 네트워크

딥러닝 학습 모델은 최근 컴퓨터 비전의 여러 영역에서 상당히 좋은 성능을 내고 있으며 많은 연구가 진행되고 있다. 본 논문에서도 딥러닝을 이용한 frame-to-frame 학습 모델을 제안한다. 제안 네트워크의 구성은 그림 1과 같은 구성을 가지고 있고 RGB-VO로 명명을 하였다.

회전 행렬과 병진 벡터에 해당하는 부분으로 병렬적으로 2개의 층으로 나뉘어져 있으며 회전 행렬을 구하는 것은 [5]에서 좀 더 어렵다고 알려져 있기 때문에 2개의 convolution 층을 더 사용하였다.

#### 2.2 데이터 세트

데이터 세트는 공공으로 제공되는 KITTI 데이터 [3]을 사용하였다. KITTI 데이터 세트는 총 00부터 21까지 22개의 시퀀스가 제공된다. 표 1은 데이터 구성을 의미하고 그림 2는 데이터의 예시를 보여준다.

Ground truth는 00부터 10까지 11개만 제공이 되기 때문에 이 11개의 시퀀스를 이용하여 학습하여 평가하였다. 학습 시퀀스로는 00부터 07까지 사용하였고 평가 시퀀스로는 08부터 10까지를 사용하였다.

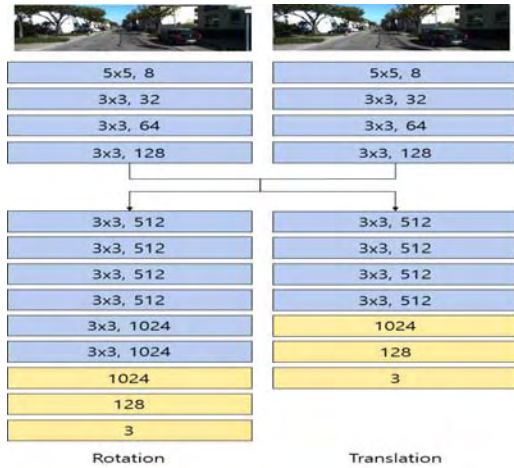


그림 1. 제안 네트워크 알고리즘. 파란 층은 convolution 필터 층(필터 사이즈, 특징 깊이)을 의미하고 노란 층은 fully-connected 층으로 벡터의 길이를 의미한다.

	시퀀스	프레임 수
학습	00	4541
	01	1101
	02	4661
	03	801
	04	271
	05	2761
	06	1101
평가	07	1101
	08	4071
	09	1591
	10	1201

표 1. KITTI 오도메트리 데이터 구성



(a) sequence 00 (b) sequence 07

그림 2. KITTI 데이터의 예시

### 2.3 학습

제안 네트워크를 학습 시, 회전 행렬은 총 9개의 값을 가지고 있다. 하지만 회전 행렬의 9개의 값은 서로 의존적인 성질이 있다. 대표적으로 회전 행렬은 symmetric한 성질이 있다. 따라서 회전 행렬을 표현하는데 9개의 값이 모두 필요하지 않고 오일러 각을 이용하여 표현하였다. 오일러 각은 3개의 값을 가지고 있으며 각 축에 대한 회전각을 의미한다. 이를 이용하여 회전 정보에 대해 학습을 진행하였다.

손실 함수로는 회전과 병진에 해당하는 부분으로 이루어져 있으며 식 (1)을 따르게 된다. 회전이 일반적으로 학습하기 어렵기 때문에  $\alpha$ 는 두 손실 함수의 정규화를 해주

는 것으로 알려져 있다 [5]. 본 논문에서는 실험적으로  $\alpha$ 를 350의 값으로 설정하였다.  $L$ 은 평균 제곱근 편차인 식 (2)을 따르게 되며  $N$ 은 총 학습 데이터 수를 의미한다.

$$\text{손실 함수} = \alpha L_{rotation} + L_{translation} \quad (1)$$

$$L = \frac{1}{N} \sum_{i=1}^N \|GT - output\|_i^2 \quad (2)$$

또한 정해진 학습 데이터를 이용한 성능을 높이기 위하여 네트워크 학습 시 3개의 프레임을 입력으로 한다. 그 후, 첫 번째와 세 번째 프레임의 관계의 카메라 예고 모션 또한 학습을 하도록 추가 하였다.

### 2.4 평가

제안 네트워크에 연속 프레임 단일 영상이 들어가 두 프레임의 오일러 각과 병진 벡터를 예측한다. 프레임 간의 결과를 귀납적으로 첫 번째 프레임에 대한 회전 행렬과 위치 정보를 각 프레임과 매칭을 한다. 즉 월드 좌표계에 대한 회전 변화량과 위치를 구하게 된다. 결과는 KITTI에서 제공하는 benchmark 코드 [3]을 이용하여 ground truth와 비교를 한다. KITTI benchmark의 결과는 회전과 병진벡터에 대한 에러이다.

### 2.4 실험 결과

실험은 비교알고리즘 VISO2-M과 SVR-VO, 제안 알고리즘 RGB-VO에 대해 하였다. KITTI benchmark 코드를 이용하여 같은 평가 데이터에 대해 실험하였다. 실험 결과는 표 2와 같은 성능을 가진다. 실험 결과는 RGB-VO가 병진 벡터 부분에서는 평균적으로 가장 좋은 성능을 나타내고 회전 행렬 부분에서는 시퀀스 10에서만 가장 좋은 성능을 가진다.

그림 3, 4, 5는 평가 시퀀스 08, 09, 10에 대해서 각각 알고리즘들을 통하여 재구성된 자동차 주행 경로를 나타낸다.

Seq	VISO2-M		SVR-VO		RGB-VO	
	Trans [%]	Rot [deg/m]	Trans [%]	Rot [deg/m]	Trans [%]	Rot [deg/m]
08	26.213	0.0247	15.42	0.0363	14.39	0.0452
09	4.09	0.0124	10.50	0.0445	9.59	0.0397
10	60.02	0.0669	21.97	0.0545	19.18	0.0438
avg	24.91	0.0266	15.02	0.0401	13.83	0.0438

표2. KITTI benchmark을 이용한 translation[%]와 rotation[deg/m]의 에러율

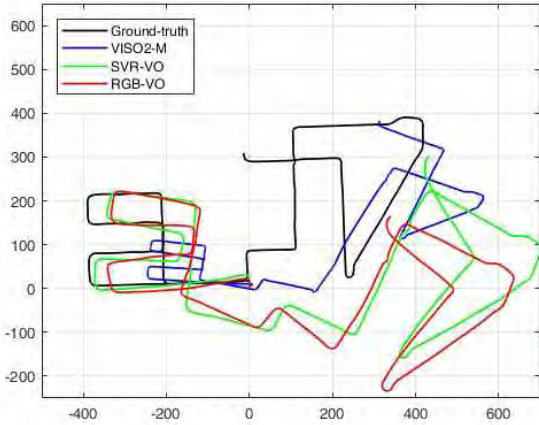


그림 3. 시퀀스 08에 대해서 재구성된 자동차 주행 경로

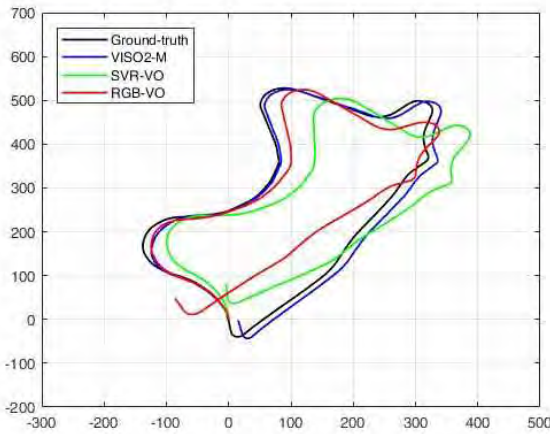


그림 4. 시퀀스 09에 대해서 재구성된 자동차 주행 경로

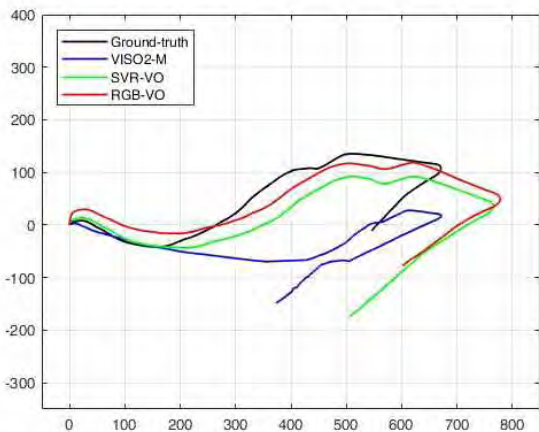


그림 5. 시퀀스 10에 대해서 재구성된 자동차 주행 경로

### 3. 결론

RGB-VO는 단일 영상만을 이용한 frame-to-frame 비주얼 오도메트리 네트워크이다. RGB-VO는 딥러닝을 이용한 학습 방법으로 기하학적 방법인 VISO2-M이 갖는 스케일 문제는 적다. 또한 다른 학습 방법인 SVR-VO는

옵티컬 플로우를 예측하여 이를 이용하기 때문에 복잡하고 옵티컬 플로우의 성능에 따라 결과가 영향을 많이 받게 된다. 하지만 RGB-VO는 단순히 RGB 도메인에서 오도메트리를 예측하는 모델이기 때문에 사용하기 간편하다. 성능은 병진 벡터를 다른 알고리즘들에 비해 잘 예측하는 것을 볼 수가 있다. 우리의 네트워크를 통하여 딥러닝을 이용한 비주얼 오도메트리의 가능성을 보여주었다. 또한 제안 네트워크에 옵티컬 플로우를 예측하여 딥러닝 네트워크를 사용한다면 좀 더 정확한 성능을 낼 수 있을 것으로 기대된다. 혹은 사전에 학습된 딥러닝 네트워크를 가져와 pre-training으로 특징을 추출하여 활용한다면 좀 더 정확하고 빠른 학습을 이뤄낼 수 있을 것이라 생각한다.

### 4. 감사의 글

이 논문은 2018년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (2016-0-00197, 스마트카 다중 센서와 딥러닝을 이용한 초정밀 내추럴 3D 뷰 생성 기술 개발)

### 참고문헌

- [1] Geiger, A., Ziegler, J., & Stiller, C. (2011, June). Stereoscan: Dense 3d reconstruction in real-time. In Intelligent Vehicles Symposium (IV), 2011 IEEE (pp. 963-968). Ieee.
- [2] Ciarruglia, T. A., Costante, G., Valigi, P., & Ricci, E. (2014). Evaluation of non-geometric methods for visual odometry. Robotics and Autonomous Systems, 62(12), 1717-1730.
- [3] Geiger, A., Lenz, P., & Urtasun, R. (2012, June). Are we ready for autonomous driving? the kitti vision benchmark suite. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 3354-3361). IEEE.
- [4] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural processing letters, 9(3), 293-300.
- [5] Alex Kendall, Matthew Grimes, and Roberto Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in Computer Vision (ICCV), 2015 IEEE International Conference on. IEEE, 2015, pp. 2938 - 2946.