

# 딥러닝 기반 시각-관성을 활용한 드론 주행기록 추정

송승연\*, 박상원\*, 김한결\*\*, 최수한\*\*

\*전남대학교 IoT인공지능융합전공

\*\*전남대학교 산업공학과

ream123@naver.com, vynx2006@gmail.com, kmr5326@naver.com, suhan9603@naver.com

## Deep Learning based Visual-Inertial Drone Odomtery Estimation

Seung-Yeon Song\*, Sang-Won Park\*, Han-Gyul Kim\*\*, Su-Han Choi\*\*

\*Major in IoT & AI convergence, Cheon-Nam National University

\*\*Dept. of Industrial System Engineering, Cheon-Nam National University

### 요 약

본 연구는 시각-관성 기반의 딥러닝 학습으로 자유분방하게 움직이는 드론의 주행기록을 정확하게 추정하는 것을 목표로 한다. 드론의 비행주행은 드론의 온보드 센서와 조정값을 이용하는 것이 일반적이다. 본 연구에서는 이 온보드 센서 데이터를 학습에 사용하여 비행주행의 위치추정을 실험하였다. 선행연구로써 DeepVO[1]를 구현하여 KITTI[3] 데이터와 Midair[4] 데이터를 비교, 분석하였다. 3D 좌표면에서의 위치 추정에 선행연구 모델의 한계가 있음을 확인하고 IMU를 Feature로써 사용하였다. 본 모델은 FlowNet[2]을 모방한 CNN 네트워크로부터 Optical Flow Feature에 IMU 데이터를 더해 RNN으로 학습을 진행하였다. 본 연구를 통해 주행기록 예측을 다소 정확히 했다고 할 수 없지만, IMU Feature를 통해 주행기록의 예측이 가능함을 볼 수 있었다. 본 연구를 통해 시각-관성 분야에서 사람의 지식이나 조정이 들어가는 센서를 융합하는 기존의 방식에서 사람의 제어가 들어가지 않는 End-to-End 방식으로 인공지능을 학습했다. 또한, 시각과 관성 데이터를 통해 주행기록을 추정할 수 있었고 시각적으로 그래프를 그려 정답과 얼마나 차이 있는지 확인해보았다.

### 1. 서론

본 연구는 시각-관성(Visual-Inertial) 기반의 딥러닝 학습으로 자유분방하게 움직이는 드론의 주행기록(Odometry)을 정확하게 추정하는 것을 목표로 한다. 주행기록은 로봇 시스템에서 자신의 상태를 기록하는 것을 말하며, 온보드 센서의 데이터를 사용하여 로봇 자체의 위치와 방향의 변화를 정확하게 추정하는 것을 목표로 한다.

드론 시스템에서도 비행 제어 구현을 위해서 드론의 상태(회전각, 각속도, 위치, 고도, 비행속도 등)를 알아야 제어 명령을 결정할 수 있는데, 일반적으로 기존 드론에서 사용하는 저가의 온보드 센서만을 이용하여 상태 추정을 하는 것이 드론의 핵심 기술 중 하나이다. 각속도계를 이용하여 회전 추정의 속도를 높이고 이와 함께 다른 센서 정보도 동시에 사용하여 IMU의 오차를 보정하는 다중 센서 융합(sensor fusion)은 드론 비행 성능 향상을 위해서는 필수적인 기술이다. 하지만 이러한 방식들은 오차에 민감하고 정밀한 주행기록 추정을 위해서는 ‘라이다

(LiDAR)’로 불리는 센서가 필요하지만, 드론 적용에 있어서 비용적, 현실적인 단점이 명확하여 최근에는 인공지능 학습을 통해 이를 보완하는 방법들이 제시되었다.

따라서 본 연구는 딥러닝 기반으로 드론 데이터에 대해 시각과 관성을 학습시켜 주행기록을 추정하며, 10% 이하의 오차를 얻는 것을 목표로 한다. 이를 위한 연구 질문은 두 가지이다. 첫째, 사람의 의도적인 조작이 들어가지 않는 End-to-End로 학습 가능한가? 둘째, 시각과 관성 데이터를 통해 드론의 주행기록을 추정할 수 있는가?

### 2. 배경 지식

#### 가. 배경 지식

##### 1) 드론 제어의 한계점

자율비행 드론이 도심의 하늘을 비행하려면 주행(또는 비행, 주변) 환경을 스스로 인식할 수 있는 기술이 필요하다. 하지만 자율비행 드론이 도심 하늘을 비행하려면 여러 제약 조건을 극복해야 한다. 대

용량 배터리를 탑재하고 컴퓨터와 고성능 센서를 설치하고, 필요한 정보를 확보한 후 복잡한 알고리즘을 적용하면 주변 환경을 충분히 인식할 수 있다. 이 방법은 드론의 무게가 무거워지고 고성능 센서를 탑재하기 어렵고 장애물과의 충돌을 피하는 제어도 쉽지 않다.

## 2) 드론 제어와 주행기록

드론 자체 센서로 환경을 인식하고 스스로 경로를 만든 후 장애물을 피해 목표 지점까지 비행하려면 환경인식·위치추정·제어 등 세 가지 핵심 기술이 반드시 구현되어야 한다. 드론 자율비행 기술은 GPS를 기반으로 하지 않는다. 그 대신 드론에 장착된 각종 센서로 환경을 인식하고, 경로를 스스로 만든 후 장애물을 피해 원하는 목표 지점까지 비행한다. 이를 위해서 주변 장애물과 드론의 상대적 위치를 인식하는 환경인식 기술, GPS 없이 스스로의 동작과 위치를 인식하는 위치 추정 기술, 정확하게 지정된 장소까지 움직이는 제어 기술이 필요하다. 이를 위해 대표적으로 쓰이는 것이 벨로다인(Velodyne) 센서, 그리고 스테레오 비전이다. 벨로다인 센서란 '라이다'로 불리는 레이저 거리 센서가 기계적으로 회전하면서 3차원 공간상의 물체 표면 좌표를 알려주는 장치다. 라이다 센서의 크기와 무게, 비행하는 드론 위에서 동적으로 작동하는 단점, 전력 문제 등으로 인해 일부에선 벨로다인 센서 대신 스테레오 비전에 의존한 환경 인식 실험을 시도하고 있다.

## 3) 딥러닝 기반의 주행기록 추정

딥러닝 기반의 방법은 비교적 최근 들어서 활발히 연구가 진행 중인 분야이다. 몇몇 기존의 파이프라인 방식을 이용한 주행기록 추정은 좋은 결과를 보여주기도 하였으나, 정확한 설계와 조정 작업에 큰 노력을 들여야 했다. 이러한 부분에서 다소 하드코딩 되었다고 할 수 있었다. 이에 여러 환경에 적용이 쉽고 조정작업이 필요 없는 End-to-end 방식의 딥러닝 모델들이 고안되었다.

### 나. 관련 연구

#### 1) Supervised Method

DeepVO[1]에서는 지도학습의 방식으로 단안 카메라의 시각데이터를 사용하여 학습하였다. 이 연구에서는 연속된 두 장의 사진에서 Optical Flow를 추출하는 FlowNet[2]의 CNN 부분을 통해 Optical flow

를 특징으로 추출하였고, 매시간의 주행기록을 추정하기 위하여 RNN이 사용되었다. 또한, End-to-end 방식으로 카메라 보정값 등 여러 조정이 필요 없이 다양한 시각데이터에 적용하여 쉽게 학습한다. 기존의 방식은 큰 특징들만 인식하여 비교하지만, DeepVO[1] 방식은 CNN을 통해 이미지의 모든 픽셀을 검사하여 특징이 두드러지지 않은 데이터에도 사용한다.

## 3. 모델 설계 및 구현

### 가. 모델 설계의 목표

연구의 목표는 드론의 영상 데이터와 IMU 센서 데이터를 End-to-End 학습하여 높은 정확도의 주행기록을 추정하는 것이다. 하지만 본 연구에서는 실제 드론을 조작하여 실험적으로 얻은 데이터가 아닌 공개된 VIO(Visual Inertial Odometry) 벤치마크 데이터셋을 활용한다.

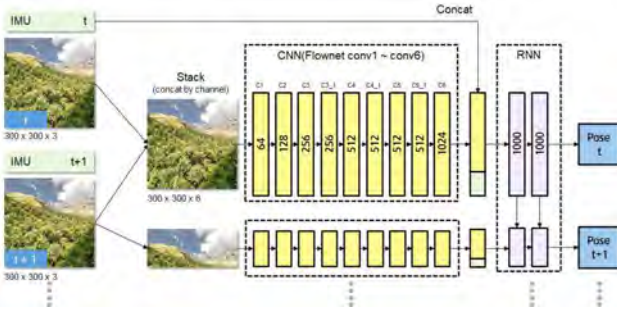
### 나. 모델 설계 및 구현

DeepVO[1]에서는 차량에서 촬영한 단안 카메라의 연속 이미지들을 CNN을 통해 Optical flow를 feature로 추출하였고, 이를 RNN을 거쳐 주행기록을 추정했다. 하지만 이 모델은 차량과 같이 수평으로 움직이는 물체는 잘 예측하지만, 드론과 같은 자유분방한 움직임을 움직이는 물체에는 적합하지 않다. 드론의 자유분방함을 생각한다면 수직 방향을 바라보고 올라가는 드론은 영상에서 지면을 많이 포함하지 않으므로 영상만으로 Optical flow를 추출하기 힘들다. 따라서 IMU를 통해 방향과 가속도를 얻어 비교적 정확한 위치를 추정할 수 있다.

연구는 DeepVO[1]의 논문을 직접 구현하는 것으로 시작한다. 그리고 DeepVO[1]는 드론이 아닌 자동차에서 얻은 KITTI[3] 데이터에 기반한 것이므로 드론의 3차원 주행에 맞는 새로운 데이터와 모델을 설계한다. 이를 위해 IMU 센서를 도입한 것으로 IMU 센서 데이터는 이미지 데이터의 주파수에 동기화하여 학습의 독립변수로 사용한다. 이를 통해 KITTI[3] 데이터, Midair[4] 데이터에 대해 DeepVO[1]에서 설계한 모델을 적용한 결과를 확인한다. 이를 통해 기존 논문에서 설계한 모델이 드론 데이터에도 성능을 보이는지 확인한다.

제안하는 모델은 [그림 1]과 같이 IMU를 학습의 독립변수로 넣었다. 드론 데이터인 Midair[4] 데이터에 설계한 모델을 적용함으로써 IMU를 Feature로

사용하였을 때 유의미한 작용을 하는지 그래프 및 수치상으로 확인한다. 또한, Midair[4] 데이터에 대해 기존 논문의 모델과 설계한 모델의 성능을 비교, 분석한다. 따라서 제안하는 모델은 드론 영상을 End-to-End 학습하여 자유분방한 움직임에 대해 더 높은 정확도의 주행기록 추정을 기대한다.



[그림 1] 본 연구에서 제안하는 모델

다. 평가방법

제안한 모델의 평가는 모델이 예측한 각 영상의 주행기록을 이용하여 여러 평가지표 및 시각화를 통해 평가한다. 평가 도구는 KITTI[3]에서 제공하는 도구를 사용하였다.

주행기록 평가에 주로 쓰이는 평가지표에는 Absolute Trajectory Error(ATE)와 Relative Pose Error(RPE)가 있다. 특히 ATE는 Visual SLAM, RPE는 Visual-Inertial Odometry에 적합하다. 두 지표는 각각 다른 장단점을 가지고 있으므로 본 연구에서 사용할 주행기록의 평가지표는 ATE와 RPE를 모두 채용하여 다음과 같이 설정하였다.

1) Absolute Trajectory Error(ATE)

ATE는 예측한 궤도를 Ground Truth의 궤도에 정렬한 후[5], 두 궤도 사이의 절대적인 거리를 계산하는 것이다. 모든 시간대에서의 발생된 오차에 RMSE를 계산한다.

이를 식으로 표현하면 :

$$ATE_{rmse} = \left( \frac{1}{n} \sum_{i=1}^n \| T_i \|^2 \right)^{\frac{1}{2}}$$

$\Delta T_i$ 는  $i$ 시간대의 위치적 오차를 의미한다. 위 식과 같이 ATE의 값은 한가지의 스칼라값으로 표현되므로, RPE보다 비교적 시스템에 대한 성능평가가 용이하다. 하지만 ATE는 주행 초기에서 발생한 오차가 끝부분에서 발생한 오차보다 ATE에 더 큰 영향을 주어 편향적인 결과를 유발할 수 있다는 단점을 가지고 있다.

2) Relative Pose Error(RPE)

RPE의 기본 개념은 주행기록 추정 시스템이 글로벌 기준을 가지고 있지 않기 때문에, 궤도를 여러 시간대로 나누어 Ground Truth와 예측값의 오차를 계산할 수 있다는 것이다[5]. 먼저, 두 궤도의 총 길이를 설정한 길이로 나누어  $k$ 개의 쌍을 만든다. 그 후, 만들어진 각각의 쌍을 서로 정렬한 후에 오차를 계산한다. RPE는 시간대별로 여러 개의 값을 가질 수 있으므로, 이 값을 분석하는 등의 활용을 통해 ATE보다 폭넓은 정보를 얻을 수 있다.

따라서 본 연구에서는 성능평가의 명확함을 위하여 이동과 회전에 대한 RPE를 각각 RMSE와 Mean Error(ME)를 사용하여 표현하였다:

$$RPE_{trans} = \left( \frac{1}{k} \sum_{i=1}^k \| \Delta t_i \|^2 \right)^{\frac{1}{2}}$$

$$RPE_{rot} = \frac{1}{k} \sum_{i=1}^k \angle (\Delta r_i)$$

4. 실험 및 분석

가. 실험 환경

1) 데이터

데이터는 Visual-Inertial Odometry Benchmark Data를 사용하였다. 그중 본 연구에서는 공개된 데이터셋인 Midair[4] Dataset, 논문에서 사용된 KITTI[3] Dataset을 사용한다.

2) 컴퓨터 사양 및 파라미터

네트워크 및 전체 학습 코드 구현에는 Pytorch를 사용하였다. 모델 훈련에는 NVIDIA Geforce RTX 2070 super GPU를 사용하여 learning rate 0.001의 파라미터로 100 epoch을 학습하였다. 또, 정확도를 높이고 훈련 시간을 줄이기 위해 CNN 부분은 사전 학습된 FlowNet[2]의 가중치를 사용하였다.

나. 실험 결과

Seq.	DeepVO		Seq.	DeepVO	
	trel(%)	rrel(°)		trel(%)	rrel(°)
03	92.408	25.158	12	126.367	70.740
04	54.080	17.971	13	124.462	105.890
06	55.492	34.612	14	161.133	109.413
07	63.729	63.768	15	138.872	70.864
10	97.031	31.278	16	149.950	74.052
mean	72.548	34.557	mean	140.156	86.1918
	KITTI			MIDAIR	

[표 1] DeepVO 모델 KITTI, Midair 적용결과

Seq.	DeepVO			Our Model		
	ATE(m)	RPE (m)	RPE (deg)	ATE(m)	RPE (m)	RPE (deg)
12	156.457	0.261	0.464	114.017	0.157	0.175
13	82.959	0.136	0.525	122.834	0.153	0.178
14	83.353	0.146	0.453	33.461	0.090	0.163
15	62.652	0.224	0.605	132.169	0.123	0.157
16	305.911	0.321	0.651	91.519	0.141	0.172
mean	138.266	0.217	0.539	98.8	0.132	0.169

MIDAIR

[표 2] Midair 데이터 DeepVO, 설계 모델 실행결과

[표 1]은 논문에서 나온 DeepVO[1] 모델에 대해 자동차 데이터인 KITTI[3]와 드론 데이터인 Midair[4] 데이터에 사용한 결과를 나타낸다.

[표 2]는 DeepVO[1]와 본 연구의 모델을 Midair[4] 데이터에 적용한 결과를 나타낸다.

다. 실험 결과 분석

1) 결과 분석

[표 1]에서 DeepVO[1]를 구현한 모델로 Translation Error, Rotation Error를 확인해본 결과, 각각의 평균이 KITTI[3] 데이터에서 성능이 좋은 것으로 확인할 수 있었다. 이 의미는 기존 논문에서 제안한 DeepVO[1] 모델이 드론 데이터에서도 어느 정도 예측은 하지만 자동차 데이터보다는 성능이 떨어진다는 것이 확인된다.

그래서 본 연구에서 제안한 모델에 대한 비교군으로 구현한 DeepVO[1] 모델로 삼았고 이 두 모델을 드론 데이터에 적용하여 그 결과를 비교하였다. [표 2]에서 Visual SLAM과 Visual Odometry에서 주로 사용되는 평가방식인 ATE, RPE를 각각 계산해본 결과, 제안한 모델에서 평균적으로 적은 오차가 나는 것을 확인하였다. 이를 통해 IMU 데이터가 유용하게 사용될 수 있는 Feature임을 확인할 수 있다.

5. 결론 및 향후과제

본 연구는 시각-관성 기반의 딥러닝 학습으로 자유분방하게 움직이는 드론의 주행기록을 정확하게 추정하는 것을 목표로 하였다. 이를 위해 선행연구로써 DeepVO[1]를 구현하여 논문에서 사용한 KITTI[3]와 Midair[4]에 적용하여 비교, 분석해보았다. 3D좌표면에서 위치추정은 Optical Flow만으로는 한계가 있음을 확인하고 IMU를 Feature로써 사용하게 되었다. 이를 통해 FlowNet[2]을 모방한 CNN의 Optical Flow Feature에 IMU 데이터를 더해 RNN

으로 학습을 진행하였다. 결과로써 RPE는 작게 나왔으나 ATE는 높게 나와 정확한 예측은 아니지만, 예측이 가능함을 볼 수 있었다.

이를 통해 연구 질문의 답을 두 가지로 내려보았다. 첫째, 기존에 VIO 분야에서 사람의 지식이나 조정이 들어가는 기존 방식에서 사람의 제어가 들어가지 않는 End-to-End 방식으로 인공지능 학습을 진행할 수 있었다. 둘째, 우수한 결과는 내지 못했지만, 시각과 관성 데이터를 통해 주행기록을 추정할 수 있었고 시각적으로 그래프를 그려 정답과 얼마나 차이 있는지 확인해보았다.

IMU Feature 이외에도 Depth, Unsupervised Feature 등을 활용해 볼 수 있으면 더 좋은 결과가 나올 것으로 생각한다. 본 연구를 통해 모델의 성능보다는 인공지능 학습의 한계를 확장했다는 것에 의미를 둔다.

[본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.]

참고문헌

[1] S. Wang, R. Clark, H. Wen and N. Trigoni, "DeepVO: Towards end-to-end visual odometry with deep Recurrent Convolutional Neural Networks," 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017, pp. 2043-2050.

[2] A. Dosovitskiy et al., "FlowNet: Learning Optical Flow with Convolutional Networks," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 2758-2766.

[3] A. Geiger, P. Lenz and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 3354-3361.

[4] M. Fonder and M. Van Droogenbroeck, "Mid-Air: A Multi-Modal Dataset for Extremely Low Altitude Drone Flights," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp. 553-562.

[5] Z. Zhang and D. Scaramuzza, "A Tutorial on Quantitative Trajectory Evaluation for Visual(-Inertial) Odometry," 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, 2018, pp. 7244-7251.