

3D Convolution과 Convolutional LSTM을 이용한 멀티모달 제스처 드론 제어

이동재, 최선웅

국민대학교 전자공학과

dj777@kookmin.ac.kr, schoi@kookmin.ac.kr

Multimodal Gesture Drone Control using 3D Convolution and Convolutional LSTM

Dong Jae Lee, Sunwoong Choi

Kookmin Univ.

요 약

본 논문은 제스처 학습을 진행하여 드론 조종을 위한 딥러닝을 진행한다. 드론 조종에는 보통 조종기를 이용하여 조종을 하지만 음성, 제스처, 웨어러블 기기 등을 이용하여 조작하는 방법들이 존재한다. 그중 제스처를 이용하여 조작하는 방법을 택하였다. 딥러닝 모델로는 Sppnet을 사용하였고 정확도는 9가지 동작에 대해서 73.37%로 확인하였다. 데이터는 IsoGD 데이터셋을 사용하였고, Training 데이터의 80퍼센트를 학습에 20퍼센트를 Test 데이터로 사용하였다. Skeleton Data와 딥러닝 모델 이용해서 12가지 동작을 만들어 10회 반복 실험 결과 평균 92.5%의 정확도를 확인하였다.

I. 서론

드론 조작에는 다양한 조작법이 존재한다. 보통 조종기를 이용하여 조종하지만 음성, 제스처, 웨어러블 기기를 이용하여 다양한 조종 방식이 존재한다.

본 논문에서는 RGB-D 카메라를 이용해 제스처를 이용해 딥러닝에 사용할 수 있는 분야에 대해 설명하고 있다. 드론을 조작하기 위해 제스처를 허용하게 되었고, 그 방법 중 제스처를 썼다. 드론 제어에는 여러 가지가 있는데 조종기, 음성, 제스처, 웨어러블 기기를 통해 하는 제어 방법들이 존재한다. 본 논문에서는 제스처를 써 딥러닝으로 학습 후 제스처로 드론을 조종하는 방법을 사용하였다. 드론을 제스처로 제어하기 위하여 사용된 기법을 소개하고자 한다. 딥러닝을 통해 모델을 학습시켜 제스처 동작을 수행한다.

II. 본론

본 논문에서는 딥러닝을 통해 제스처를 학습하여 드론 조작 방법을 제안한다. 모델은 Sppnet을 이용하여 학습을 진행하여 테스트하였다.

드론 조작에는 다양한 기법들이 존재한다. 스마트 글로브와 한 손 조종기와 같이 기기를 통해서 조작하는 방법들을 자주 사용한다. 스마트 글로브는 AHRS(Attitude Heading Reference System) 센서를 활용한 직관적인 제스처 컨트롤이 가능하다. 블루투스 통신을 통하여 기기를 연동하여 데이터를 불러와 동작을 수행한다. 다른 방법으로는 한 손 조종기도 존재한다. 한 손 조종기는 직접 드론이 이동할 목적지를 포인팅하는 방식을 사용하고, 드론과 조종기 사이에는 보이지 않는 선이 연결된 것과 같이 작동한다. 조종기의 가속도 및 자이로 센서를 통해서 조종 메커니즘을 통해 조종기와 드론 간의 거리 산출을 통해서 드론을 제어하게 된다. 음성인식을 통한 제어 방식도 존재한다. 마이크로 통해 음성인식 시 사용자의 음성을 저장하고 음성인식 라이브러리인 Google Assistant를 이용하여 처리된 음성으로도 드론 제어가 가능하다. 하지만 본 논문에서는 카메라를 이용

해서 제스처를 인식하여 드론 조작이 가능한 방법을 딥러닝을 이용해서 찾게 되었다. Sppnet은 기존의 CNN에서 Spatial Pyramid pooling layer를 포함시킨 모델이다. Convolution3D와 Convolutional LSTM을 사용하고 Batch normalization과 MaxPooling을 통해서 학습을 진행하고, 본 논문에서는 RGB와 Depth 영상 두 가지를 사용하여 마지막에 Concatenate를 진행 후 Fully connected를 진행하게 된다.

모델 내부 구조로는 Convolution3D의 Filter size = [64, 128, 256, 256], Kernel size = [3, 3, 3], MaxPooling size = [(1, 2, 2), (2, 2, 2)]로 진행하였고, Convolutional LSTM을 사용하였다. Spp layer(Spatial pyramid pooling layer)의 부분은 본 논문에서는 4개의 피라미드를 사용하게 되는데, [7x7, 4x4, 2x2, 1x1]의 4가지 영역으로 나뉘어서 학습을 진행하였다. RGB와 Depth 두 가지 Input으로 학습을 진행하고 각각 Spp layer의 값을 Concat을 진행하여 마지막 Fully connected를 진행하였다.

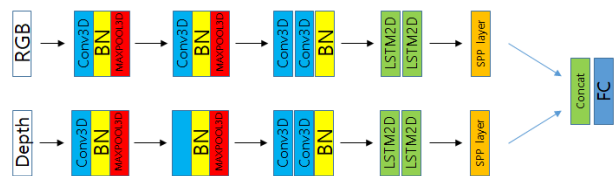


그림 1 제안된 네트워크 구조

입력 데이터는 IsoGD 데이터 셋을 사용하였다. 데이터 전처리 과정에서는 전체 데이터 평균을 15개로 나누어서 15장의 데이터를 추출하여 입력 데이터로 사용하였다. 총 9가지 동작에 대해서 학습을 진행하였다.

Azure Kinect를 이용하여 드론 조작 방법을 제안한다. Kinect를 통해 Skeleton Data를 얻기 위해서는 Microsoft에서 제공하는 SDK를 활용한다. 두 손의 움직임을 통해 드론을 제어하도록 구현하였다. 드론 제어는 모드를 나누어 설계하게 되었다.

드론 제어 모드는 두 가지로 구현되었다. 대기모드와 조종모드로 나뉘는데, 대기모드는 드론을 조종하기 전과 후에 조종에 대한 오작동이 발생하지 않도록 넣어준 모드이다. 조종모드는 조종에 필요한 손동작으로 시작된다. (그림 2)는 드론을 조작하기 위한 준비 자세이다. 양손의 팔꿈치가 직각이 되도록 들어 팔을 앞으로 뻗어주는 제스처가 대기모드에서 드론을 조작하는 포즈이다.

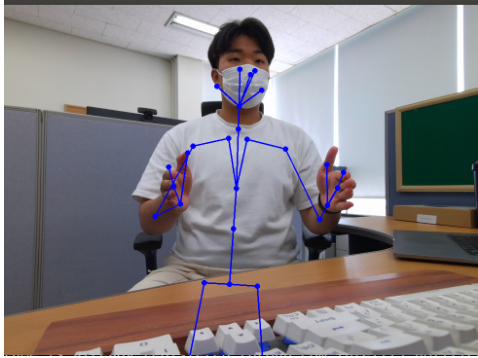


그림 2 Drone Takeoff pose

드론 시뮬레이터로는 PX4 Autopilot(그림 3)의 프로그램을 사용하였다. 시뮬레이터를 이용해 MAVSDK의 명령어를 통해 드론을 조작한다.

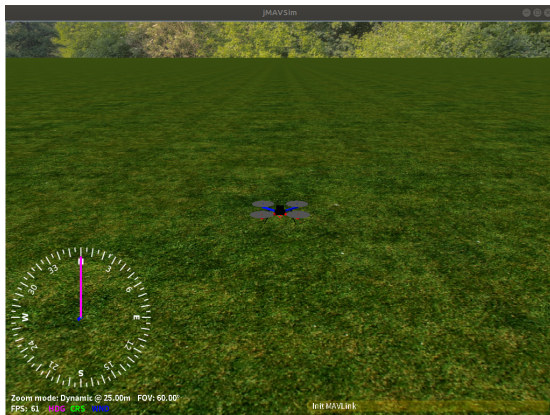


그림 3 드론 시뮬레이터(PX4 Autopilot)

드론 조작으로는 4가지 레버가 존재하는데 드론 조종기와 같이 팔을 이용해서 8가지 방향을 조작할 수 있다. 조종기의 레버를 움직이는 것처럼 두 팔을 이용해서 조작하면 된다. Check, Arming, Takeoff, Go Up, Go Down, Pin Left, Pin Right, Go Straight, Go Back, Go Left, Go Right, Land 등으로 12가지(표 1) 제스처 동작을 구현하였다.

표 1 명령어 리스트

조종 동작	Check	Takeoff	Pitch Up	Roll Left	Yaw Left	Throttle Up
	Arming	Land	Pitch Down	Roll Right	Yaw Right	Throttle Down

카메라로 드론을 제어할 때 문제점이 발생하게 되는데, 카메라 안에 여러 사람이 존재 할 경우가 가장 큰 문제점이다. 조종자 한 명을 집어주어야 드론 제어가 가능하므로 그 문제점을 가장 가까운 인물의 데이터를 먼저

전달하여 제어하는 방법을 사용하였다. 만약 가장 가까운 조종자가 자리를 이탈하여 화면에서 사라지면 다시 카메라와 가까운 사람이 조종자가 되는 방법을 사용하였다.

III. 결론

본 논문에서는 학습 데이터로는 IsoGD 오픈 데이터를 사용하였다. Sppnet 모델을 사용하여 12 epoch를 학습하였고, Train Accuracy는 99.71%이고, Test Accuracy는 73.37%이다. Train과 Test set의 비율은 Train 80%, Test 20%로 나뉘었다. 실험 환경은 구글 코랩을 이용하여 학습하였고, Tensorflow와 Keras를 사용하여 학습을 진행하였다. Check 동작에 사용이 되었고, 나머지 11가지 동작은 Skeleton Data를 이용해서 드론 조종기와 같이 설정해주었다.

실험 환경은 Ubuntu 18.04버전으로 진행하였다. 드론 시뮬레이터로는 PX4 Autopilot을 사용하였고, 드론 조작에 대한 명령어 조작은 MAVSDK를 이용하여 명령어를 전달하였다. 프로그래밍 언어는 Python을 사용하였다. 딥러닝 학습에는 Tensorflow를 사용하였다.

표 2 테스트 결과

명령어	실험 자수									
	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차
Check	T	T	T	F	T	T	T	T	T	T
Arming	T	T	T	T	T	T	T	T	T	T
Takeoff	T	T	T	T	T	T	T	T	F, T	T
Go Up	T	T	T	T	T	T	T	F	F	T
Go Down	T	T	T	T	T	T	T	T	T	T
Pin Left	T	T	T	T	T	T	T	T	T	T
Pin Right	T	T	T	T	T	T	T	T	T	T
Go Straight	T	T	T	T	T	T	T	T	F	T
Go Back	T	T	T	T	T	T	T	F	T	T
Go Left	T	T	T	T	T	T	T	T	F	T
Go Right	T	T	T	T	T	T	T	T	F	T
Land	T	T	T	T	T	T	T	F	T	T
True 합계	12	12	12	11	12	12	12	10	7	12
False 합계	0	0	0	1	0	0	0	2	6	0

True : 112 / False : 9 / Total : 121

12가지 동작에 대하여 10회 반복 실험 결과 평균 92.5%의 제스처 인식 정확도를 보였다. 명령어 Check 와 같이 간단한 제스처를 이용해서 통합 연구 인터페이스에서 기체 상태나 각종 상태 창을 열 수 있는 제스처로 이용할 수도 있을 것 같다. 드론 조종 중 사용자가 편하게 쉴 수도 있는 모드인 Kruze 모드도 추가할 계획이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부의 재원으로 한국연구재단, 무인이동체원천기술개발사업단의 지원을 받아 무인이동체원천기술개발사업을 통해 수행되었음(No.2020M3C1C1A01084900, 2021R1F1A1062285)

참 고 문 헌

- [1] Chalearn LAP IsoGD Database, <http://www.cbsr.ia.ac.cn/users/jwan/database/isogd.html>
- [2] Kaiming He, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1904 -1916, SEP. 2015.
- [3] Guangming Zhu, "Multimodal Gesture Recognition Using 3-D Convolution and Convolutional LSTM", IEEE Access, pp. 4517 - 4524, March 2017.