

OpenPose를 활용한 음성인식기반 드론제어 촬영시스템

조유진, 김세현, 권예림, 정순호
부경대학교 컴퓨터공학과

yougin2123@pukyong.ac.kr, 3hyun@pukyong.ac.kr, moonctp24@pukyong.ac.kr,
snow@pknu.ac.kr

Speech-Recognition Drone Camera System using OpenPose

Yu-Jin Cho, Se-Hyun Kim, Ye-Rim Kwon, Soon-Ho Jung
Dept. of Computer Engineering, Pukyong University

요 약

최근 드론과 1인 미디어 시장의 성장으로, 영상 촬영 분야에서의 드론 산업이 활발하게 발전되고 있다. 본 논문에서는 딥러닝 기반 다중 객체 인식 기술인 Openpose를 활용하여 인물촬영을 위한 음성 인식 드론 제어 시스템을 제안한다. 해당 시스템은 자연어 처리된 음성명령어를 통해 드론이 각 촬영 객체에 대한 회전, 초점변화 등 실제 영상촬영기법에 사용되는 다수의 동작을 수행할 수 있도록 한다. 최종적으로 96.2%의 정확도로 음성명령에 따라 동작을 수행하는 것을 확인할 수 있다. 이는 누구나 전문적 지식이나 경험 없이 음성만으로 쉽게 드론을 제어할 수 있을 것으로 기대된다.

1. 서론

최근 드론과 1인 미디어 시장의 성장으로, 영상 촬영 분야에서의 드론 산업이 활발하게 발전되고 있다. 또한, 누구나 쉽게 드론을 조종할 수 있는 음성을 이용한 드론 모터 제어[1], 음성 인식을 기반으로 드론을 통한 재난 현장 인명 구조[2] 등 음성을 통한 드론 제어 방법이 많은 분야에서 활용되고 주목 받고 있다.

한편, 딥러닝 기반 다중 객체 인식 기술은 작업자 위치 추정 및 예측하는 협동 로봇 시스템[3], 도심지 자율 주행을 위한 실시간 다중객체인식방법[4] 등 다양한 산업에서 활용되고 있다. 이 중에서 OpenPose는 사진이나 영상에서 실시간으로 사람들의 몸, 손, 그리고 얼굴의 특징점을 추출할 수 있는 기술을 말한다. 헬스나 동작 판단과 같은 분야에서 이러한 OpenPose를 활용 기술을 적용한 사례[5][6]는 많지만, 드론과 함께 적용한 사례는 거의 찾기가 어렵다.

본 논문에서는 OpenPose를 활용한 인물 인식 및 다양한 영상촬영 기법을 제공하는 음성인식기반 드론제어 촬영시스템을 구현하고자 한다.

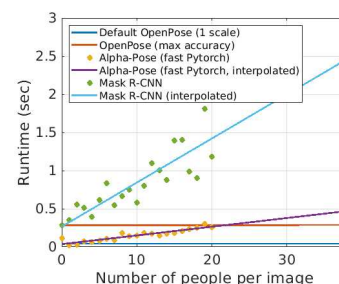
이어지는 2장에서는 시스템의 이론적 배경을 소개

하고, 3장에서 음성인식기반 영상촬영 드론 제어 시스템의 구성을 소개, 4장에서 구현 및 실험, 5장에서 결론을 언급하며 마무리하겠다.

2. 이론적배경

2-1. OpenPose

OpenPose는 사진이나 영상에서 실시간으로 사람들의 몸, 손 그리고 얼굴의 특징점을 추출할 수 있는 라이브러리를 말한다. 모든 사람의 관절을 추정하고, 각 관절의 위치를 이은 다음, 각각에 해당하는 사람의 관절 위치로 재생성하는 Bottom-Up 방식을 사용하기 때문에 반복적인 처리 없이 성능을 향상시킬 수 있다.



(그림 1) RunTime Analysis Graph.[7]

(그림 1)은 여러 Human Detection 알고리즘의 특정 이미지 속 사람의 인원수와 runtime의 관계를 그래프로 나타낸 것이다. 이에 따르면 AlphaPose, Mask R-CNN과 달리 OpenPose는 이미지 내 사람 수가 많아져도 일정한 성능을 보인다[7]는 장점을 가진다.

3. 음성인식 기반 드론제어 촬영 시스템

3-1. 시스템 구성

해당 시스템은 다양한 촬영동작을 음성명령 한 문장에 축약하기 위해 쿼트콧터의 기본 동작인 Yaw, Roll, Throttle, Pitch의 속도조절과 함께 ‘인물 대상 관점의 움직임’을 고안했다.

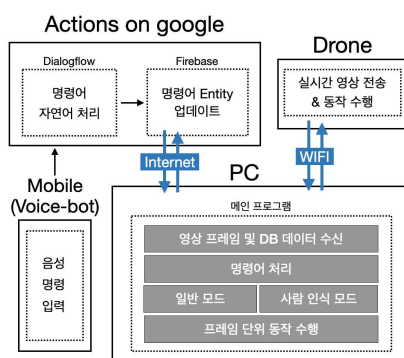
(가) “오른쪽으로 움직여”

“앞으로 가면서 아래로 가줘”

(나) “2번 사람 트래킹 해주세요”

“첫번째 사람 오른쪽으로 회전해”

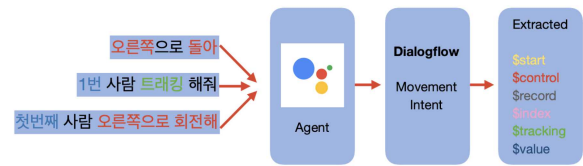
다시 말해, 본 시스템은 드론 자체의 기본 동작으로 예문 (가)와 같은 단순 명령과 더불어, 객체기반 트래킹을 통해 예문 (나)와 같은 복합 명령을 자연어 처리해낸다. 즉 관절인식 딥러닝모델 Openpose를 활용한 Tracking mode를 통해, 음성 명령으로 영상촬영과 같은 다양한 형태의 드론 동작을 구현하게 된다는 것이다.



(그림 2) 전체 시스템 구성도

(그림 2)는 본 논문에서 설계한 영상촬영 드론제어 시스템의 전체적인 모습이다. 먼저 사용자가 기존의 Voice-bot 시스템을 활용하여 음성 명령어를 입력하면, NLU 엔진을 통해 해당 명령이 5가지 Entity로 분류되고, 실시간 DB는 이를 업로드하게 된다. 이와 동시에 메인 프로그램에서는 드론의 실시간 영상을 프레임 단위로 송출함과 동시에 실시간 명령에 따른 드론 동작을 수행한다.

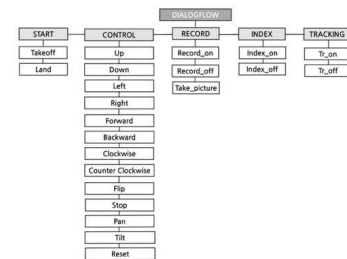
3-2. 드론 제어를 위한 시스템 설계



(그림 3) Dialogflow를 활용한 자연어 처리 과정

음성인식 드론 조종을 위해서는 Google 개발자용 플랫폼 ‘Actions on google’을 활용하며, Voice-bot을 통해 입력받은 음성명령을 자연어 처리한다. 따라서 (그림 3)에서와 같이 ‘Dialogflow’를 사용해 동작 Intent를 뽑아내고 키워드 기능에 따라 Entity를 추출한 후 실시간 데이터베이스에 등록된다. 이러한 자연어 처리된 명령어 Intent는 언제든지 사용자의 요구에 따라서 추가적으로 학습시킬 수 있는 피드백형 서비스를 구축할 수 있다.

(a) 명령어 처리



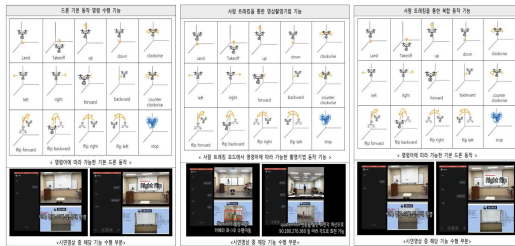
(그림 4) DB 내 명령어 집합을 구성할 수 있는 Entity 형태

(그림 4)는 PC의 메인 프로그램이 DB로부터 얻을 수 있는 5가지 형태의 명령어 집합 모습이다. 본 시스템은 이에 따라 프레임 단위로 명령어 선후관계를 파악하고, 그 순서대로 각각의 수행 함수를 실행한다. 또한, 회전명령어나 촬영명령의 경우, 한 번 명령을 받아온 다음 해당 동작을 완성할 때까지 반복 실행하지 않도록 DB를 초기화해주는 처리를 수행한다.

(b) 프레임 단위 동작 수행

프레임 단위의 드론통신과정에서는 음성명령에 따라 각각의 드론 기본 동작 속도(Yaw, Roll, Pitch, Throttle)를 변화할 수 있도록 한다. 이는 (그림 5)에서 보인 바와 같이 0과 음, 양, 즉 세 형태의 속도를 구성하여 총 81가지의 기본동작을 한 문장으로 수행할 수 있도록 했다. 또한 드론에서 실시간 프레임임을 받아들일 때, 복잡한 영상처리를 줄일 수 있도록 프레임을 일부 스킵하며 압축한다. 이 때 촬영모드

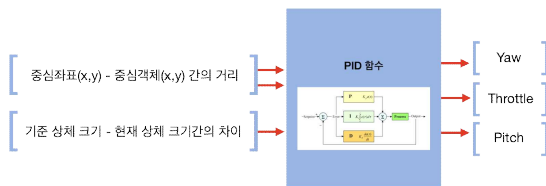
일 경우, 원본 프레임 자체에 대한 저장을 할 수 있도록 하여 화질엔 피해가 없도록 한다.



(그림 5) 수행 가능한 단일 동작 메뉴얼 모습

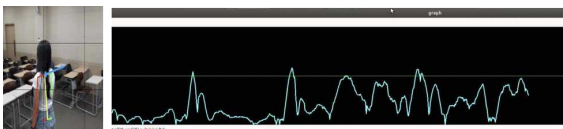
(c) PID 제어를 활용한 Tracking 기능

더하여 기본동작을 증가하는 여러 관점에서의 움직임을 위해서 PID 제어를 활용한다. PID 제어는 드론 개발에서 중요한 기본원리로, 보통 가속도 자이로 센서를 이용한 모터 출력량 조절에 쓰인다. 본 시스템은 해당 원리를 활용하여 중심좌표와 객체간의 '오차'라고 할 수 있는 거리가 발생할 경우, 이를 상쇄시키는 속도로 속력값을 변경한다. 같은 원리로, 프레임 내 객체의 상체크기의 변화를 감지하며 동일한 거리를 유지한다.



(그림 6) PID 제어를 활용한 속도 자동 조절 과정

각각의 오차값을 입력변수로 넣은 PID는 (그림 6)과 같이 해당 오차를 줄이기 위한 특정 속도를 출력해내며, 3차원 움직임을 제어하여 객체를 트래킹할 수 있다. 기본적으로 드론은 프레임 단위의 즉각적인 속도 변화를 유지해야 하기 때문에 비례상수 P를 위주로 활용하였지만, 거리 제어를 위한 상체 크기 값에 대해서는 세밀한 조정을 위해 PID를 모두 사용했다.



(그림 7) 제어를 위한 상수 값 변경 과정 중 한 그래프의 모습

제어를 위한 상수값 변경 과정 중 일부는 (그림 7)으로 확인할 수 있다. 이러한 PID 상수값 설정을 통해 순간 x축 차이가 최대 100pixel 전후를, 순간 y축 차이가 최대 30pixel 전후를 넘지 않는 안정된

드론 주행을 진행할 수 있다.

(d) 회전 및 거리 제어

우선 세부적인 회전 제어를 위해서는 드론 내부에 존재하는 센서로 얻어지는 사원수를 활용한다. 이 때, 객체에 대한 회전 각을 계산하며 '뒤', '오른쪽', '왼쪽'과 같은 방향 문자열 '재활용'을 통해 드론이 90, 180, 270도 등 다양한 각도로 회전할 수 있다.

또한, tracking mode를 위해 프레임 내 객체크기 유지로 거리제어를 구현하였는데, 이 때 상체길이와 어깨길이의 비율이 2배가 넘으면, 옆모습이라고 판단하게 된다. 이어서 옆모습의 경우, 어깨길이 측정 오류를 대비해 상체 가로 길이에서 세로 길이로 기준 길이를 변경시켜준다. 결과적으로 이는 (그림 8)과 같이 다양한 촬영위치에서의 거리제어가 가능하도록 할 수 있다.



(그림 8) 회전/거리 제어를 위한 인물 모습

마지막으로, 객체 인식 오류를 대비하여 관절의 비율을 활용했다. 몸을 비틀거나 숙여서 일부 관절이 측정되지 않을 경우를 대비하여, 어깨길리와 상체길리를 목하부와 코를 통해 가늠할 수 있도록 한다. 타겟의 크기가 너무 크다면 안전모드를 발생시켜, 시스템이 모든 동작을 중지할 수 있도록 하여 혹시 모를 촬영사고에 대비할 수 있도록 할 수 있다.

4. 시스템 정확도 테스트 및 결과

본 시스템의 드론 동작 제어 정확도를 실험하기 위해 드론의 동작을 크게 안전제어, 회전 동작, 촬영 기법 수행 동작, 기본 동작으로 나누어 회전동작은 60회, 나머지는 각각 30회씩 실내에서 진행하였고 결과는 (표 1)과 같다.

안전제어의 경우, 프레임에서 인식된 사람의 중심좌표 위치가 너무 아래에 있을 때 드론이 머리 위로 지나쳐버리는 경우가 있었다. Safe mode의 조건을 만족할 만큼 크기가 충분하지 않은 사람은 지정된 위험 크기에 도달하지 못했기 때문에 멈추지 않고 그대로 머리위로 지나가는 경우가 발생한 것이다. 이는 해당 객체에 대한 위험상황이 아니기 때문에

안전제어에 대한 문제가 없다고 판단할 수 있다. 회전 모드에서는 사용한 텔로 드론 자체의 문제로, 복합 명령 사용 시 배터리 소모량이 커져 180도 이상의 회전은 칼나의 각도를 측정하지 못해 반 바퀴를 추가로 더 돌게 되어 아주 적은 확률로 오류가 발생하였다.

구분	안전제어	회전동작	촬영기법	기본동작
성공 횟수	$\frac{26}{30}$	$\frac{59}{60}$	$\frac{30}{30}$	$\frac{30}{30}$
정확도	86.7%	98.3%	100%	100%

(표 1) 드론 동작 수행 정확도 실험 결과

결과적으로 (표 1)과 같이 안전제어는 30회 중 26회, 회전동작은 60회 중 59회, 촬영 기법 수행 동작은 30회 중 30회, 기본 동작은 30회 중 30회 성공을 하였다. 각각의 정확도를 퍼센트로 변환하면 각각 86.7%, 98.3%, 100%, 100%로 평균 96.2%의 정확도로 음성명령에 따라 동작을 수행하는 것을 확인할 수 있었다.

5. 결론 및 기대효과

전 세계적으로 드론 시장의 규모는 점점 증가하고 있으며, 특히 영상촬영과 항공촬영을 위한 드론 사용의 비중이 증가하고 있다. 본 논문에서는 OpenPose를 활용하여 인물 인식 및 다양한 영상촬영 기법을 제공하는 음성인식기반 드론제어 촬영시스템을 제안하였다. 해당 시스템을 통해 드론 비전문가, 그리고 드론을 조종할 신체적 여건이 되지 않는 사람들도 전문적 지식이나 경험 없이 음성만으로 쉽게 드론을 제어할 수 있을 것으로 기대된다. 나아가 재난 상황 혹은 엔터테인먼트 분야 등 다양한 분야에서 발전할 수 있을 것이다.

참고문헌

- [1] 김민수, 김덕래, 김진수, 김선형. “음성을 이용한 드론 모터 제어”. 한국정보기술학회. 하계공동학술대회 및 대학생논문경진대회. 한국정보기술학회, 2019.
 - [2] Yuki Yamazaki, Masaya Tamaki, Chinthaka Premachandra, C. J. Perera, Sagara Sumathipala and B. H. Sudantha “Victim Detection Using UAV with On-board Voice Recognition System” 2019 Third IEEE International Conference on Robotic Computing. IEEE, 2019.
 - [3] 임지웅, 임성수. CNN 물체 인식 알고리즘을 이용한 협동 로봇 시스템 내 작업자 위치 추정 및 예측. 대한기계학회 춘추학술대회, 1268-1272, 2019
 - [4] 박상배, 김정하. 도심지 자율주행을 위한 머신러닝 기반의 실시간 다 객체 인식 방법. 제어로봇시스템학회 논문지, 26(6), 499-505, 2020
 - [5] 한재혁, 최지훈, 박구만. Openpose 라이브러리를 활용한 헬스 자세교정 프로그램. 한국 방송 미디어 공학회 학술발표대회 논문집, 48-48, 2018
 - [6] 유혜림, 이봉환. 감시 영상을 활용한 OpenPose 기반 아동 학대 판단시스템. 한국정보통신학회논문지, 23(3), 282-290, 2019
 - [7] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, “Realtime multi-Person 2D Pose Estimation using Part Affinity Fields,” in Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, p7, 2016.
- [본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.]