

# 2D Human Pose Estimation 기술 분석

차진혁, 정승원  
 동국대학교 멀티미디어공학과  
 e-mail : [ckwlsgur20@naver.com](mailto:ckwlsgur20@naver.com)

## A Study on 2D Human Pose Estimation Techniques

Jin-Hyuck Cha and Seung-Won Jung  
 Dept. of Multimedia Engineering, Dongguk University

### 요 약

딥러닝 기술의 발전에 따라, 딥러닝을 Human Pose Estimation 에 적용하는 연구가 활발하게 수행되고 있다. 본 논문에서는 딥러닝을 이용한 여러 기술 중 가장 활발하게 사용이 되고 있는 Open Pose 와 Deeper Cut 기술의 특성을 분석한다.

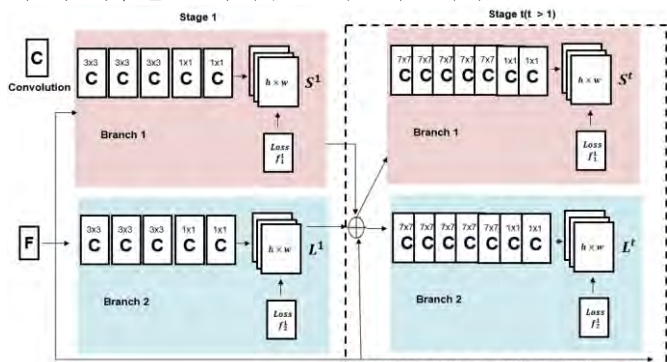
### 1. 서론

최근 딥러닝을 이용한 Human Pose Estimation 기술에 대한 연구가 활발히 진행되고 있다. 기존에 키넥트를 이용한 Human Pose Estimation 기술은 적외선을 이용하기에 야외에서는 정확도가 떨어졌고, Marker 를 이용한 추정 기술이 있으나, 기술을 사용하기 위해 내부의 공간과 비싼 장비가 필요했다. 하지만 딥러닝을 이용한 Human Pose Estimation 기술은 값싼 카메라 하나로도 정확도가 높은 결과물을 만들 수 있다. 이런 장점을 가진 기술을 2 가지 조사해 보았다.

### 2. Open Pose

첫 번째 기술은 Open Pose [1]이다. 2D 이미지를 넣어주게 되면 사람의 Keypoint 를 찾아준다. 이런 Keypoint 를 찾아 주어 최종적으로 사람의 자세를 추정할 수 있도록 각 키포인트를 연결해 준다.

입력으로 들어온 이미지를 이미 학습된 VGG19 의 앞쪽의 10 개의 레이어를 이용해 feature map 을 추출한 후 네트워크에 넣어주게 된다. 네트워크는 part confidence map 과 part affinity fields 두 브랜치로 나뉜다. Part confidence map 은 body part 를 찾아주는 네트워크고, Part affinity fields 는 여러 사람이 있는 이미지 일 시 각 사람을 분리해주는 네트워크이다.



(그림 1) Open Pose 기술의 네트워크 구조

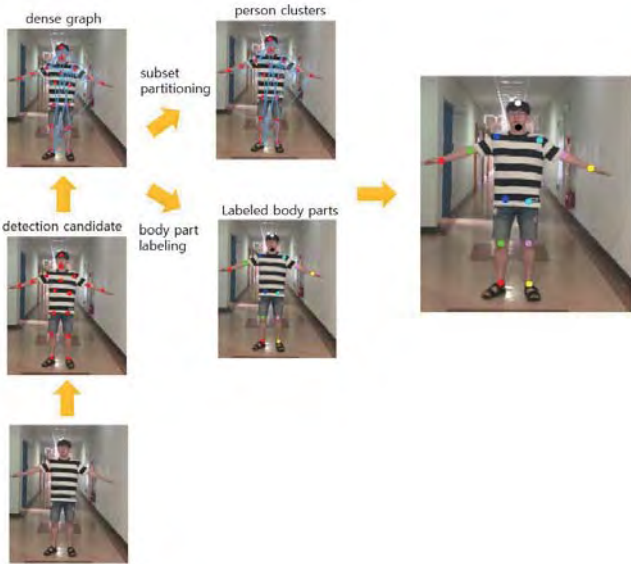
이렇게 두 브랜치 네트워크를 이용해 여러 사람이 있는 이미지에 대해 Keypoint 를 찾아 낼 수 있다. OpenPose 는 real time multi-person detection 에 좋은 성능을 내며, 19 명이 있는 비디오에서 8.8fps 의 성능을 낸다.



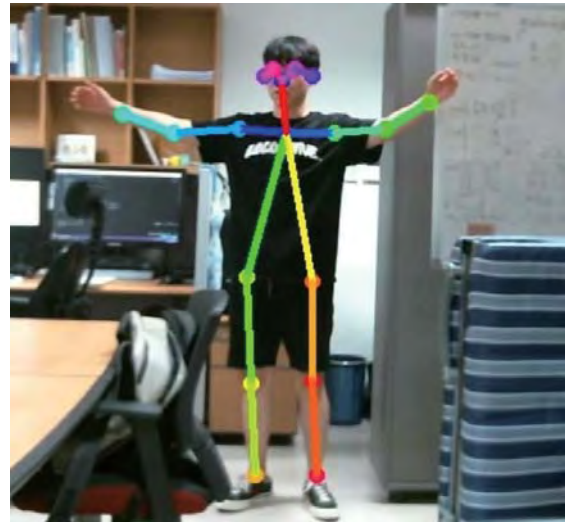
(그림 2) OpenPose 기술의 결과 예시

### 3. Deeper Cut

두 번째 기술은 Deeper Cut [2]이다. Deeper Cut 은 사람이 있는 단일 이미지 한장이 인풋이며, 인풋 이미지에서 body part detector 를 이용해 body part candidate 를 구한다. Body part detector 는 Res-Net 을 사용한 Adapted Fast R-CNN 이다. Body part candidate 를 모두 연결해 포화 그래프를 만든 뒤, subset partitioning 을 통해 각각의 사람을 clustering 하고, body part labeling 을 통해 각 body part 의 부위를 명확히 지정해준다. 이후 clustering 된 정보와 labeling 된 정보를 취합해 Keypoint 를 찾은 이미지를 결과로 내준다.



(그림 3) Deeper Cut 기술의 흐름도



(그림 5) 비디오가 인풋일 때 Open Pose 결과

#### 4. 비교

한 명의 사람만이 있는 이미지를 인풋으로 넣었을 때 결과물을 비교해 보았다.



(그림 4) 단일 이미지 결과물 비교

비교 결과 단일 이미지 한장만을 넣었을 때 Open Pose 보다 Deeper Cut 이 더 정확히 body part detection 이 되었다. 특히나 Open Pose 의 경우 사람의 자세가 역동적일 수록 팔 다리와 같은 말단 부위가 detection 이 잘 안되는 현상을 보였다. (그림 4)의 테니스 하는 사람의 자세를 보면, 굉장히 역동적인데, Deeper Cut 의 경우 굉장히 잘 측정 되었지만, Open Pose 의 경우 몸체는 잘 측정되었지만, 말단 부위로 갈수록 detection 이 잘 되지 않았다.

Open Pose 의 경우 비디오가 인풋으로 들어갈 수도 있어 비디오를 인풋으로 넣어보았다.

Open Pose 는 비디오를 input 으로 넣었을 때가 단일 이미지를 넣었을 때 보다 결과물이 훨씬 좋았고, 한 명만 있는 경우에는 30FPS(Frame per Second)까지 렌더링이 가능했다.

#### 5. 결론

단일 이미지 만을 이용해 Keypoint 를 찾는 것이면 Deeper Cut 이 Open Pose 에 비해 더 정확히 찾지만, 속도가 느려 실시간으로 사용이 불가능 하다. 반면 Open Pose 는 단일 이미지에 대해서는 결과물이 Deeper Cut 보다 좋지 않지만 인풋을 비디오로 넣었을 때는 단일 이미지를 인풋으로 넣었을 때 보다 더 좋았다. 또한 속도 또한 빨라 실시간으로 사용 가능한 기술이기에 실제 사용하기에는 Open Pose 가 더 실용적일 것이라 생각된다.

#### 참고문헌

- [1] Z Cao, T Simon, SE Wei and Y Sheikh. "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields". In CVPR, 2017
- [2] E Insafutdinov, L Pishchulin, B Andres, "DeeperCut: A Deeper, Stronger and Faster Multi-person Pose Estimation Model". In ECCV, 2016