# 추종로봇을 위한 대상 추적 및 중심 추정 방법

이준구, 이지은, 원혜민, 오지용
한국전자통신연구원
{leejg01679, jieun.lee, hyemin_won, jiyongoh}@etri.re.kr

# Target tracking and centroid estimation method for following robots

Joon-Goo Lee, Jieun Lee, Hye-min Won, Jiyong Oh
Electronics and Telecommunications Research Institute (ETRI)

요 약

최근 로봇의 발전과 함께 다양한 산업 분야에 로봇의 사용이 대중화되고 있으며, 특히 운송을 위한 모바일 추종로봇에 대한 연구가 활발하다. 이를 위한 요소 기술로 소개된 딥러닝 기반의 최신 추적 기술들은 대상의 클래스 정보를 사용하지 않아 유사한 특징의 다른 대상과 식별 번호가 혼동되는 문제가 발생하며, 깊이 정보를 추출하기 위한 박스 기반의 중심 및 최근접 포인트 추정은 사용자의 자세에 따라 잘못된 포인트를 추정하여 로봇으로부터 다양한 오류를 야기할 수 있다. 본 논문에서는 skeleton 정보를 이용하여 추적 대상의 클래스 정보를 보완하고, 추적 대상의 중심점을 보다 강인하게 추출하는 방법을 소개한다.

## Ⅰ. 서 론

최근 로봇의 발전과 함께 다양한 산업 분야에 로봇의 사용이 대중화되고 있다. 특히 모바일 로봇은 무거운 중량 및 부피가 큰 물건들을 옮기기 위해 사용되고 있으며, 이를 위해 사용자를 추종하는 로봇에 대한 연구가 활발하다. 이를 위한 요소 기술로 대상 객체를 추적하는 추적 기술과 2차원 이미지에서 추출된 객체를 실세계의 3차원 정보로 변환하기 위한 deprojection 기술이 필요하다. 일반적으로 객체 추적을 위해서는 딥러닝을 기반으로 하는 객체 추적기술이 많이 사용되며, 실세계 좌표로 투영하기 위해 깊이 정보를 제공하는 depth 카메라를 함께 사용한다.

하지만 최근 성능이 높은 딥러닝 기반의 객체 추적 기술은 클래스 정보를 사용하지 않기 때문에 유사한 특징 정보와 혼동되어 추적에 실패할 수 있다. 또한 대상의 깊이 정보를 추출하기 위한 방법으로 박스 기반 중심점 추정 및 박스 기반 최근접 포인트 추정은 자세에 따라 대상이 아닌 곳을 추정하여 로봇이 오작동하는 문제가 발생한다.

본 논문에서는 이러한 문제를 해결하기 위해 pose estimation의 skeleton 정보를 사용하여 대상 클래스 추정을 보완하였으며, skeleton 랜드마크의 중심 포인트를 이용하여 자세에 강인한 중심 추정법을 제안하였다.

## Ⅱ. 본론

본 논문에서는 딥러닝 기반의 객체 추적 기술과 대상의 중심 추출을 위해 그림 1과 같이 기존의 딥러닝 추적 방법에 pose estimation의 skeleton 정보를 함께 고려하는 방법을 소개한다. 먼저 최신 딥러닝 기반의 객체 추적 기술은 동일한 두 개의 CNN(Convolution Neural Network) stream을 사용하여 추적 대상 이미지와 검색할 전체 이미지의 특징을 추출한 후 전체 이미지의 특징에서 대상 이미지의 특징과 관련 있는 영역을 검출한 후 추적하는 Siamese 기반의 방법[1]과 이미지를 패치로 분할 하여 포지션 정보와 임베딩한 후 encoder, decoder 및 self-attention을 적용하여 추적하는 Transformer 기반의 방법[2]이 주로 소개되고 있다. 하지만 이러한 방법은 객체의 유무 또는 유사도를 판단하는 binary classifier와 객체의 영역 박스를 추정하는 regression을 헤더로 사용하여 실제 추적하는 객체가 어떤 객체인지를 표현하는 정보가 없다. 즉, 초기에 설정되어 추적 중인 사람으로부터 추출된 특징 정보만을 사용하기 때문에 그림 2의 왼쪽과 같이 사람과 크기와 형상이 유사한 다른 대상과 혼동되어 추적이 실패하는 경우가 발생한다.

이러한 문제를 해결하기 위해 기존 딥러닝 기반의 추적 결과에 간단하게

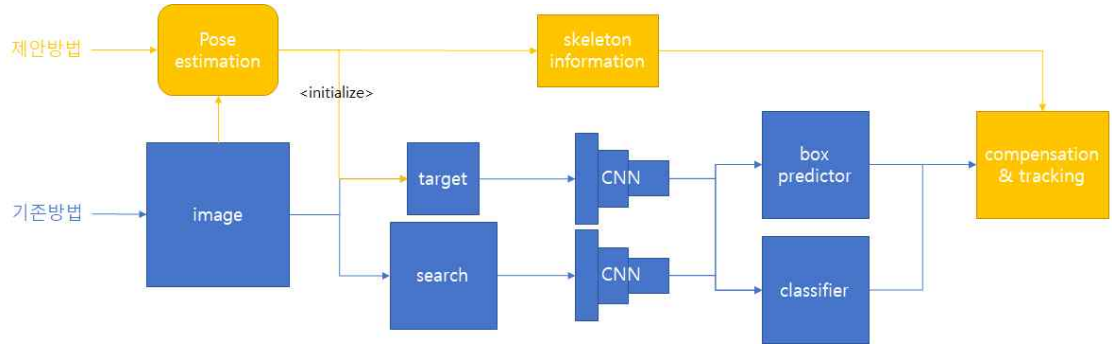


그림 1. 제안 방법의 구조도

그림 2. 클래스 정보의 유무에 따른 추적 대상 혼동의 예

그림 3. 추적 대상의 중심 포인트 설정 방법

결합하여 사용할 수 있는 pose estimation의 skeleton 정보[3, 4]를 함께 고려한다. 기존의 딥러닝 기반 추적의 결과는 클래스의 정보 없이 초기 선택된 대상을 박스 영역으로 추출하기 때문에 대상이 사람인지 아닌지 구분할 수 없다. 그림 2의 오른쪽과 같이 skeleton 정보는 사람인 경우에만 estimation 되기 때문에 기존 추적기가 추적한 영역에 skeleton 정보가 포함되어 있으면 이를 사람으로 추적하고, 해당 영역에 skeleton 정보가 존재하지 않으면 추적 대상이 잘못된 것으로 판단하여 재추적하도록 설정한다.

대상을 바르게 추적한 후 로봇이 해당 사람을 추종하기 위해서는 영상 좌표를 실세계 3차원 좌표로 변환해야 하며, 이를 위해서는 추적 대상의 거리를 추출해야 한다. 실험에 사용된 카메라는 RGBD 센서인 Stereo Labs의 ZED2 카메라를 사용하였으며, RGBD 센서는 RGB 영상과 정합된 Depth map 또는 Point Cloud를 제공하기 때문에 대상의 특정 포인트를 할당하면 해당 좌표의 거리 또는 실세계의 3차원 좌표를 추출할 수 있다. 2차원 영상에서 추적 대상의 특정 포인트를 할당하기 위한 방법으로 박스를 기반으로 중심 영역을 이용하는 방법과 박스 내에서 거리가 가장 근접한 포인트를 설정하는 방법이 있다. 하지만 이러한 방법은 그림 3의 왼쪽과 같이 사용자의 자세에 따라 중심이 허공을 가리키거나 발끝 또는 어깨와 같이 depth map의 hole이 발생할 수 있는 위치를 가리켜 로봇이 오작동 하는 경우가 발생한다.

이러한 문제를 해결하기 위해 앞서 추정된 skeleton의 랜드마크를 이용하여 중심을 결정한다. 해당 알고리즘은 MS-COCO 데이터셋을 기반으로 학습하여 그림 3의 오른쪽과 같이 17개의 skeleton 랜드마크 포인트가 추출되며 이중 11번과 12번 랜드마크 포인트는 자세와 상관없이 항상 사람의 골반을 가리키게 된다. 따라서 두 좌표 사이의 중심점 또는 두 좌표의 평균 값을 이용해 해당 좌표의 depth 정보 또는 실세계의 3차원 좌표를 추출한다. 추출한 depth 정보와 camera intrinsics 정보를 이용해 실세계 3차원 좌표로 deprojection한 후 모바일 로봇의 선각속도를 계산하면 로봇이 대상을 바르게 추종할 수 있다.

## Ⅲ. 결론

본 논문에서는 사용자 추종로봇을 위해 클래스 정보를 고려한 추적 방법 및 depth 정보 추출을 위한 중심 추정 방법을 소개하였다. 기존의 딥러닝 기반 추적 방법에서 부재한 클래스 정보를 pose estimation의 skeleton 정보를 함께 고려하여 다른 클래스와 혼동되는 경우를 개선했으며, 자세의 영향을 받는 중심 추정 방법에서 skeleton의 랜드마크 포인트를 이용하여 다양한 자세에도 강인한 객체의 중심을 추정하는 방법을 제안하였다. 실험에서 제안한 방법은 기존의 방법에서 발생하는 문제점들을 잘 보완하였다.

## ACKNOWLEDGMENT

본 논문은 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임(“학연협력플랫폼구축 시범사업” RS-2023-00304776).

## 참 고 문 헌

[1] Z. Zhang et al, “Ocean: Object-aware Anchor-free Tracking,” Computer Vision-ECCV 2020: 16th European Conference, Part XXI 16, pp.771-187, 2020.

[2] Ye, Botao, et al, "Joint feature learning and relation modeling for tracking: A one-stream framework," European conference on computer vision. pp.341-357, 2022.

[3] Cao, Zhe, et al, "Realtime multi-person 2d pose estimation using part affinity fields." Proceedings of the IEEE conference on computer vision and pattern recognition. pp.7291-7299, 2017.

[4] Xiao, Bin, et al, "Simple baselines for human pose estimation and tracking." Proceedings of the European conference on computer vision (ECCV). pp.466-481, 2018.