

클래스 정보를 고려한 삼 네트워크 기반의 추적 방법

이준구, 오지용, 남승우

한국전자통신연구원

leejg01679@etri.re.kr, jiyongoh@etri.re.kr, swnam@etri.re.kr

A Tracking Method Based on Siamese Network with Class Information

Joon-Goo Lee, Jiyong Oh, Seung-Woo Nam

Electronics and Telecommunications Research Institute (ETRI)

요약

본 논문은 휴먼 추종을 위한 삼(Siamese) 네트워크 기반 추적기에 객체의 클래스 정보를 함께 고려하는 방법을 제안한다. 삼 네트워크 기반의 추적기는 클래스의 구분 없이 선택된 대상으로부터 추출된 특징만을 이용하기 때문에 추적 중인 클래스의 객체와 유사한 특징을 갖는 다른 클래스의 객체를 혼동하기 쉽다. 따라서 객체의 클래스 정보를 추출할 수 있는 검출기를 결합하여 추적 대상의 초기화를 수행하고 추적 중인 객체가 동일한 클래스인지 주기적으로 비교한다. 실험에서 제안하는 방법은 기존의 방법보다 유사한 클래스가 존재하는 경우에 대해 더 강인하게 작동하였다.

I. 서론

영상기반 추적 기술은 생활/방범, 교통 감시, 물류 로봇 등의 다양한 산업 분야에서 응용되어 필요성이 점차 증가하고 있으며, 특히 물류 로봇 분야에서는 휴먼 추종과 관련된 기술 수요가 증가하고 있다. 최근 딥러닝 기술의 사용이 보편화 되면서 추적 기술에도 다양하게 접목되고 있으며, 삼(Siamese) 네트워크 기반 추적 기술은 높은 성능과 빠른 처리 속도로 인해 효율이 매우 우수하다[1]. Zhang은 삼 네트워크 기반의 추적기에 기존의 검출 목적인 백본 네트워크 대신 추적에 적합한 백본 네트워크를 제안하였으며, 변형 가능한 합성곱(deformable convolution)을 적용하여 객체의 경계 상자를 더 정확하게 검출하였다[2-3]. 하지만 삼 네트워크 기반의 추적기는 실시간으로 입력되는 영상에 대한 초기화 방법이 적절하지 않고, 사용자가 초기에 설정한 대상으로부터 추출된 특징만을 고려하기 때문에 추적 중 대상이 혼동되는 경우가 발생한다.

본 논문에서는 객체의 클래스 정보를 추출할 수 있는 검출기를 삼 네트워크 기반의 추적기와 결합하여 추적 대상의 초기화를 수행하고, 추적 중인 객체가 동일한 클래스인지 추가적으로 비교하여 추적 성공률을 높이는 방법에 대하여 제안한다.

II. 본론

제안하는 방법은 그림 1과 같이 기존의 추적기에 검출기를 결합한 형태로 구성된다. 기존의 추적기는 이미지를 입력받아 추적하고자 하는 대상과 검색 이미지를 생성하고, 동일한 가중치(weights)를 공유하는 CNN(Convolution Neural Network)을 통해 특징을 추출하여 경계 상자 예측 및 특징 분류를 수행한다. 결합된 검출기는 입력 이미지에 대해 검출을 수행하고, 검출 결과로 출력된 객체의 경계 상자와 클래스 정보를 추적 결과와 함께 고려하여 초기화 및 추적을 수행한다. 여기서 추적기는 Ocean 추적기를 사용하였으며[3], 검출기는 YOLOv4-tiny를 사람만을 대상으로 검출하는 1-클래스 검출기로 변형하여 사용하였다[4].

먼저 기존의 추적기 초기화 방법은 저장된 영상의 첫 프레임에 대하여 마우스 드래그 또는 미리 설정된 좌표를 이용해 추적 대상의 경계 상자를 입력하였다. 하지만 실시간 영상에서는 초기 프레임의 경계 상자를 사전에 추출할 수 없으며, 마우스 드래그 조작시 객체의 움직임으로 인해 정확한 경계 상자를 설정하기 어렵다. 따라서 그림 2와 같이 검출기를 이용해 실시간으로 입력되는 프레임의 객체들을 추출하여 추적 대상의 상자를 선택하거나 화면의 특정 위치에 일정 시간 이상 존재하는 대상의 상자를 자

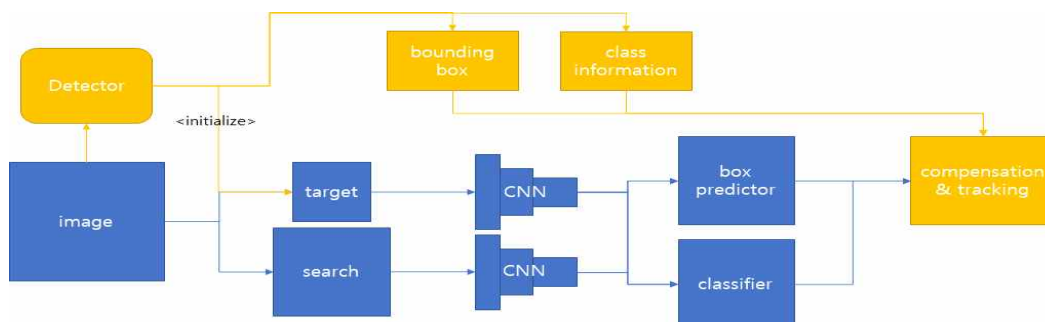


그림 1. 제안 방법의 구조

동으로 선택하여 추적 대상을 초기화한다. 여기서 녹색 상자는 검출기가 검출한 객체의 경계 상자이고, person은 클래스명, (354, 235)는 경계 상자의 중심이다. 1.64는 사전에 설정한 추적 시작 점과의 유클리드 거리로 임계값 10 미만으로 가까워지면 추적을 수행하도록 설계하였다.



그림 2. 초기화 방법

제안하는 방법의 성능 비교를 위해 1분 단위의 영상 5개를 선정하여 테스트를 수행하였으며, 기존의 방법은 5개 영상 중 2가지 영상에 대하여 추적에 실패하였다. 기존의 삼 네트워크 기반의 추적기는 클래스 정보를 고려하지 않고 초기 설정된 대상의 특징만을 사용하기 때문에 형태나 컬러가 비슷한 다른 클래스와 혼동되기 쉽다. 그림 3은 이러한 추적 오류의 예를 보여주고 있으며, 여기서 녹색 경계 상자는 추적기가 정상적으로 추적 중인 대상을 나타낸다. (a), (c)는 초기 설정한 사람을 대상으로 추적하고 있었으나 카메라와 대상이 움직이면서 (b), (d)와 같이 유사한 형태 또는 유사한 컬러의 다른 클래스의 객체로 추적 대상이 변경되었으며, 추적 신뢰도 역시 90% 이상으로 잘못된 대상을 계속 정상적으로 추적하였다.

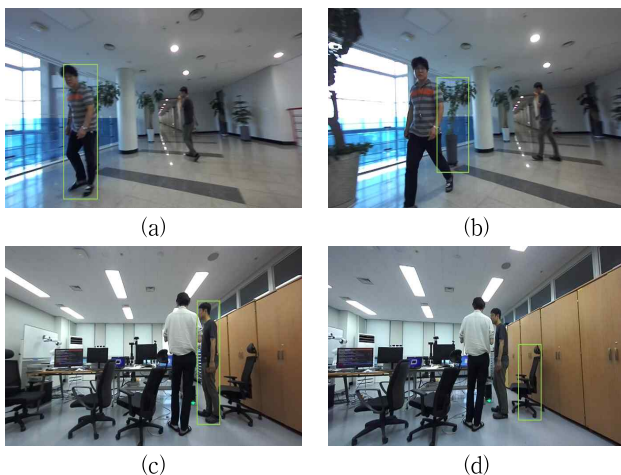


그림 3. 추적 오류 예

제안하는 방법은 초기에 검출기로부터 설정된 클래스 정보와 현재 프레임의 검출기로부터 검출된 경계 상자를 함께 고려하여 문제점을 보완하였으며, 테스트 영상 모두 추적에 성공하였다. 그림 4는 기존의 방법으로 추적에 실패한 영상에 대하여 제안하는 방법을 수행한 결과를 보여준다. (a), (c)의 녹색 경계 상자는 추적기를 통해 추적 중인 대상이며, 파란색 경계 상자는 검출기를 통해 검출된 객체의 경계 상자들 중에 IoU가 가장 가까운 객체이다. 초기에 추적 대상으로 설정된 클래스와 현재 검출되어 매칭된 경계 상자의 클래스가 같기 때문에 정상으로 판단하여 추적을 진행한다. (b), (d)의 빨간색 경계 상자는 추적기를 통해 추적 중인 대상이고, 녹색

색 경계 상자는 제안하는 방법을 통해 보정된 결과이다. 초기에 추적 대상으로 설정된 클래스와 현재 검출되어 매칭된 경계 상자의 클래스가 다르기 때문에 비정상적으로 판단하고, 이전 프레임에서 정상으로 추적하던 대상과 현재 프레임에서 검출된 객체들 중에 IoU가 가장 높은 경계 상자를 선택하여 재 초기화 후 다시 추적을 수행하였다.

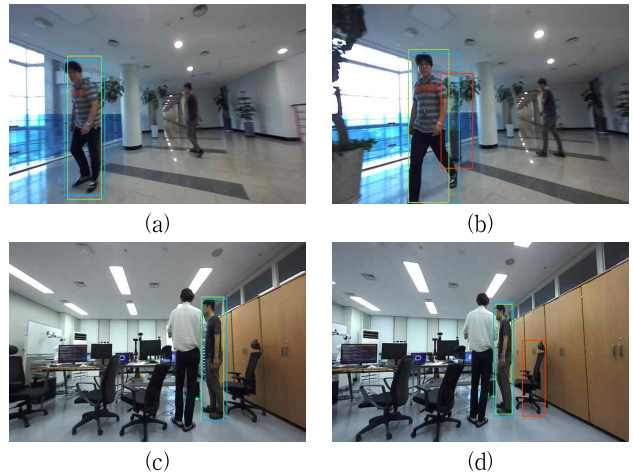


그림 4. 제안 방법 결과 예

III. 결론

본 논문에서는 기존의 삼 네트워크 기반의 추적기와 검출기를 결합하여 다른 클래스와 혼동되는 상황에서도 강한 추적 방법을 제안하였다. 또한 추적기의 초기화에도 검출기를 사용하여 실시간으로 입력되는 영상에 초기화가 가능하도록 하였다. 실험에서 제안하는 방법은 기존의 방법보다 우수한 성능을 나타내었다.

추후 RGB만 사용하는 기존의 추적 방법에 Depth 정보를 함께 고려하여 동일한 클래스 내에서 혼동되는 경우에도 강인하게 추적할 수 있는 방법을 연구할 예정이다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZD1130, 지능제어기반 스마트기계 및 로봇 기술 개발]

참 고 문 헌

- [1] Berinetto L., et al., "Fully-convolutional siamese networks for object tracking," European conference on computer vision, pp. 850-865, 2016.
- [2] Z. Zhang, et al., "Deeper and wider Siameses networks for real-time visual tracking," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4591-4600, 2019.
- [3] Z. Zhang, et al., "Ocean: Object-aware Anchor-free Tracking," Proceedings of 16th European Conference on Computer Vision, pp. 771-787, 2020.
- [4] Bochkovskiy A., et al., "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.