

딥러닝 기반 실시간 다중 객체 추적 시스템

*김경훈, **강석주
서강대학교 전자공학과

*godgang@sogang.ac.kr, **sukjukang@sogang.ac.kr

Deep Learning-based Real-Time Multi-object Tracking

*Kyunghun Kim and *Junho Heo and *Sukju Kang
*Sogang University

요 약

최근 객체 추적 분야에서의 딥러닝 활용도가 증가하고 있다. 대표적인 딥러닝 기반 객체 추적 시스템은 객체 검출 알고리즘과 추적 알고리즘의 연쇄된 형태로 구성된 형태이다. 따라서 추적 시스템의 성능은 객체 검출기의 성능에 매우 의존적이라는 한계점이 존재한다. 본 논문에서는 위와 같은 검출-추적의 직관적인 구조로 구성된 시스템에서 딥러닝 기반 검출기에 대한 의존성을 낮추고 또한 추적기의 동작 속도를 증가시키는 방법을 제안한다.

1. 서론

비디오 영상에서의 자동 물체인식 및 추적 기술은 자율주행, 스포츠 경기 분석, 무인감시 및 범죄예방 등 다양한 분야에서 활용되고 있으며, 국내외에서 관련 기술의 개발이 크게 급증하고 있다. 하지만 기존의 영상 내 여러 객체를 동시에 추적하는 다중 객체 추적(MOT) 또는 다중 대상 추적(MTT) 기술은 무궁무진한 활용도에 비해 부족한 성능을 보인다. 컴퓨터 비전 분야에서 MOT의 구성은 크게 세가지로서 먼저 객체들의 위치 할당, 객체의 정체성(ID) 분류 그리고 각 객체의 궤적 생성으로 나뉜다[1].

객체 추적이 어려운 이유에는 여러가지가 있는데 대표적으로 탐지기의 성능이 좋지 못해서 탐지를 못하는 경우와 다른 객체에 가려져 탐지가 불가능한 경우가 있다. 다중 객체의 추적이 이루어지기 위해서는 먼저 영상 내 존재하는 객체들의 검출과정이 선행되어야 한다. 따라서 객체의 검출 정보를 토대로 추적을 수행하는 알고리즘은 검출기의 정확성에 크게 좌우된다. 초기의 다중 객체 추적 기술은 이전 프레임에서 탐지한 객체들의 정보를 가지고 객체들의 다음 위치 및 속도를 예측하고, 그 예측 값과 현재 프레임에서의 객체 탐지 값을 비교하여 두 값 간의 최소차이들을 매칭시켜주는 방식으로 진행되었다. 그러나 객체 탐지가 제대로 되지 않을 경우 이러한 객체 추적 방식은 객체들의 정확한 매칭이 어려워 추적의 정확성이 낮아질 수밖에 없다. 이러한 추적 문제를 해결하기 위해 객체 탐지기의 성능을 높이기 위한 연구와 객체가 가려져도, 즉 객체간 가려짐이 발생하여도 추적이 유지되기 위한 연구가 진행되고 있다. 이 외에도 객체 탐지 및 추적 속도를 개선하는 연구들이 진행되고 있다.

최근 Deep Sort[3] 논문에서 알고리즘 구조 단순화를 통한 속도 개선과 가려짐에 대해 강건한 성능을 보이는 연구가 제안되었다. Deep Sort 알고리즘은 검출된 객체를 기반으로 추적 시 기존 추적 알고리즘과 딥러닝 기반 알고리즘을 함께 사용하도록 고안되었다.이 딥러닝 네트워크를 사용해 이전

프레임에서 추적하던 객체의 특징 정보를 추출하여 저장하고, 이후 프레임에서 새롭게 탐지된 객체의 특징 정보가 이전의 객체 특징 정보와 같다고 판단 될 경우 이 둘을 동일 객체로 매칭한다. 본 논문에서는 기존 Deep Sort 알고리즘의 객체 특징 정보 비교 과정을 변형하여 성능을 향상시키는 방법을 제안한다.



그림 1. Deep Sort 를 사용한 객체 추적 영상

2. 본론

기존의 Deep Sort 알고리즘은 매 프레임마다 객체의 위치, 속도 및 특징 정보를 추출하기 위해 검출기와 딥러닝 기반 특징 추출기를 사용하기 때문에 전체 동작시간에서 가장 큰 비율을 차지한다. 따라서 해당 부분의 처리 과정을 줄이는 방향으로 추적기의 동작 속도를 향상하는 실험을 그림 2 와 같이 진행하였다.

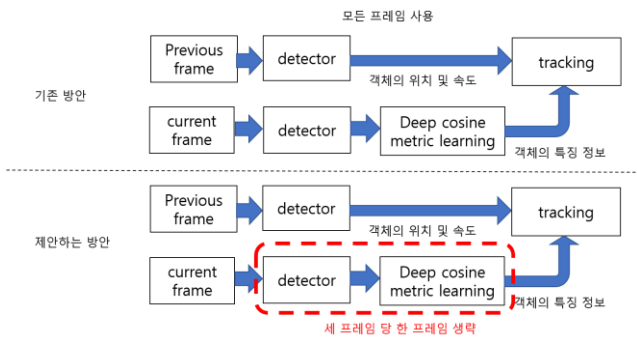


그림 2. 기존 방안과 제안하는 방안

본 실험에서는 객체 탐지기로 YOLO-v3 모델을 사용하여 다중 객체 탐지를 수행하였고 Deep Sort 알고리즘을 사용하여 다중 객체 추적 시스템을 구현하였다. 실험 데이터는 MOT16 을 사용하였고 MOT challenge benchmark 사이트에서 제공하는 다중 객체 추적 성능 평가 tool 을 사용하여 추적 정확도 성능을 평가하였다(그림 3).

solution	video	MOTA	MOTP	AVG-FPS(CPU)
original	MOT16-04	34.3	79.2	3.77
	MOT16-09	45.7	74.5	4.62
	MOT16-10	31.4	75.6	4.40
	MOT16-11	45	79.2	4.46
	summary	35.9	78.2	4.31
our	MOT16-04	34.5	79.4	12.81
	MOT16-09	46.5	74.8	14.14
	MOT16-10	30.6	75.6	14.31
	MOT16-11	46	79.1	14.41
	summary	36.1	78.4	13.92

그림 3. MOT challenge benchmark tool 을 통한 다중 객체 추적 성능 평가 변경 전(위)과 변경 후(아래)

첫번째 실험은 기존 YOLO-v3 모델과 기존 Deep Sort 알고리즘을 적용하여 실험하였다. 정확도는 평균 35.9 점을 얻었고 평균 fps 는 4.31 로 측정되었다. 두번째 실험은 탐지 프레임을 줄이고 추출된 특징 정보의 저장 기간을 늘려 성능을 평가하였다. 실험 결과, 제안한 방법이 정확도 성능에서 미세하게 높은 점수를 얻었고 속도 성능이 약 3 배정도 향상된 결과를 얻게 되었다. 즉 모든 프레임에서 객체 검출을 하지 않아도 추적기의 정확도가 떨어지지 않고 동작 속도가 크게 개선됨을 실험을 통해 확인하였다.



그림 4. 폐색 발생 시 추적 성능 결과

3. 결론

본 논문에서는 기존 객체 검출-추적 기반의 연쇄된 형태의 알고리즘에서 객체 검출 정도를 조절하여 추적 성능을 향상시키는 방법을 제안한다. 매 프레임마다 객체 검출을 수행하는 대신 3 개 프레임 간격으로 검출 빈도수를 조절하여 전체 추적기의 동작 시간을 실시간 처리가 가능하도록 개선하였고, 이전 프레임에서 탐지한 객체의 feature 정보 저장 기간을 늘려 오랜 프레임기간 폐색이 발생하여도 추적 정확성을 높이는 방안을 제안한다.

본 연구는 2018 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 기초연구사업의 지원을 받아 수행되었으며 (No. 2018R1D1A1B07048421), 또한 행정안전부 스마트 도로조명 플랫폼 개발 및 실증 연구 개발사업의 연구비지원(과제번호 19PQWO-B153369-01)에 의해 수행되었습니다.

참고문헌

- [1] Luo, Wenhan & Xing, Junliang & Milan, Anton & Zhang, Xiaoqing & Liu, Wei & Zhao, Xiaowei & Kim, Tae-Kyun. (2017). Multiple Object Tracking: A Literature Review.
- [2] Bewley, Alex et al. "Simple Online and Realtime Tracking." 2016 IEEE International Conference on Image Processing (ICIP) (2016): n. pag. Crossref. Web.
- [3] N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 3645-3649
- [4] N. Wojke and A. Bewley, "Deep Cosine Metric Learning for Person Re-identification," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, 2018, pp. 748-756.
- [5] A. Milan, L. Leal-Taix'e, I. Reid, S. Roth, and K. Schindler, "Mot16: A benchmark for multi-object tracking," arXiv preprint arXiv:1603.00831, 2016.
- [6] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," EURASIP J. Image Video Process, vol. 2008, 2008.