

다중 랜덤 워커 기반 추적기를 활용한 동영상에서의 객체 검출 기법

임경선, 김한울, 김창수

고려대학교

{kslim, hanulkim, cskim}@mcl.korea.ac.kr

Video object detection algorithm with multiple random walkers based tracker

Kyungsun Lim Han-Ul Kim Chang-Su Kim

Korea University

요약

본 논문에서는 동영상에서 객체를 자동 검출하는 기법을 제안한다. 제안하는 기법은 정지 영상에서 객체를 검출하는 기법과 동영상에서 객체를 추적하는 기법을 동시에 수행하여 동영상에서 객체를 검출한다. 매 프레임 검출기는 학습된 종류의 객체들을 검출하고 추적기는 이전 프레임에서 검출되었던 객체를 추적한다. 검출기가 검출한 결과와 추적기가 추적한 결과를 매칭하고, 겹치는 결과와 그렇지 않은 결과에 대해 각각 다른 검사를 수행하여 신뢰도 있는 결과를 도출한다. 실험 결과를 통해 제안하는 기법이 기존 검출 기법에 비해 우수한 성능을 보임을 확인한다.

1. 서론

영상에서 객체를 검출하는 기술은 영상을 분석하고 이해하는데 있어 핵심적인 주제로 최근 연구가 활발하다. 특히 인공지능영향을 이용한 객체 검출 기법[1][2]들은 객체 검출 분야의 큰 성능 향상을 이끌어 다양한 분야의 기초기술로 사용되고 있다. 하지만 이들 기법의 경우 정지 영상에 대해 개발되어 무인 감시 시스템 그리고 자동 주행 시스템과 같이 동영상에서의 분석을 필요로 하는 시스템에 적합하지 않다. 따라서 본 논문에서는 객체 검출 기법[2]과 추적 기법[3]을 조합하여 동영상에서 효과적인 객체 검출 및 추적이 가능하도록 하는 알고리즘을 제안한다.

본 논문에서 이전 프레임 객체에 대한 추적을 수행하기 위해 다중 랜덤 워커 기반의 객체 추적 기법[3]을 사용한다. 이와 독립적으로 인공 신경망 기반의 객체 검출 기법[2]을 통해 현재 프레임에서 객체 검출을 수행한다. 추적결과와 검출결과가 주어졌을 때 검출기가 검출한 결과와 추적기가 추적한 결과를 매칭하고, 겹치는 결과와 그렇지 않은 결과에 대해 각각 다른 검사를 수행하여 신뢰도 있는 결과를 도출한다. 20개의 실험 영상에 대한 실험 결과는, 제안하는 기법이 기존 검출 기법보다 우수한 성능을 보임을 확인한다.

본 논문의 구성은 다음과 같다. 2장에서 제안하는 기법을 통해 동영상 객체를 검출 및 추적하는 과정을 설명한다. 3장에서는 실험을 통해 제안기법의 우수함을 확인한다. 마지막으로 4장에서 본 논문에 대한 결론을 맺는다.

2. 제안하는 기법

2.1 객체 검출 및 추적 결과 획득

효과적인 동영상 객체 검출을 위해 제안하는 기법은 객체 검출 알고리즘[2]과 객체 추적 알고리즘[3]을 결합한다. 매 프레임 객체 검출과 이전 프레임에 존재했던 객체에 대한 추적을 각각 수행한다. 매번 독립적으로 작동하는 검출기와 다르게 추적기는 이전 프레임에서 학습된 분류기로 현재 프레임에서 객체를 추적한다. 분류기는 객체 박스내의 배경과 객체를 효과적으로 분리하기 위해서 박스를 다수의 블록으로 나누어 다중 랜덤 워커 시뮬레이션을 수행한다. 각 블록에 객체가 존재할 확률을 해당 블록의 특징 벡터에 가중치로 곱하여 배경의 영향을 억제한다. 이는 객체 추적 벤치마크 영상에서 기존 기법과 비교하여 높은 정확도를 보인다.

2.2 객체 검출 및 추적 결과 매칭

검출된 객체 박스와 추적된 객체 박스간의 IoU(Intersection over Union) 비율을 계산하여 동일한 객체를 나타내는 경우를 가려낸다. 검출된 박스와 추적된 박스간의 IoU 비율이 0.5 이상인 경우, 매칭된 박스로 정의하고 두 박스는 동일 객체를 포함 하는 것으로 간주한다. 따라서 매칭 이후 모든 검출 결과와 추적 결과는, 매칭된 결과, 매칭되지 않은 검출 결과, 매칭되지 않은 추적 결과 세 가지 경우로 분류 할 수 있으며, 각각 다른 접근 방법을 취해 현재 프레임에서의 객체를 판단한다.

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업의 연구결과로 수행되었음(IITP-2016-R2720-16-0007).



(a) Faster R-CNN (b) 제안하는 기법
 그림1. Faster R-CNN[1] 과의 정성적 비교

	precision	recall	F1 score
Faster R-CNN(θ_d)	0.919	0.473	0.599
Faster R-CNN(θ_t)	0.532	0.785	0.632
Proposed	0.867	0.691	0.741

표 1. Faster R-CNN과의 성능 비교

대한 추적기를 갱신한다.

2.3 매칭된 객체 박스의 처리

일반적으로 검출 스코어가 높은 박스일수록 객체를 정확하게 표현하는 경향을 보인다. 또한 검출된 박스는 높은 검출 스코어를 보인 박스로 구성하였으므로 정확하게 객체를 표현하고 있다고 판단할 수 있다. 따라서 매칭된 객체 박스의 경우 객체의 최종 위치는 검출기의 결과를 따른다.

2.4 매칭되지 않은 추적된 객체 박스의 처리

추적된 박스와 일치하는 검출 박스가 없는 경우는 객체가 사라졌거나 검출기가 객체를 놓친 것으로 생각할 수 있다. 따라서 추적된 박스가 여전히 객체를 포함하고 있는지에 대한 검사가 필요하다. 검출기를 통해 추적된 박스에 대한 검출 스코어와 재조정된 박스 위치를 얻는다. 이 때, 검출 스코어가 실험적으로 구한 문턱 값 $\theta_t = 0.5$ 이상이면 추적된 박스를 객체 박스로 판단한다. 이후 식 (1)과 같이 추적된 박스와 재조정된 박스를 조합하여 객체의 최종 위치를 결정한다.

$$b = x + f(x_{score})(x' - x) \quad (1)$$

$$f(s) = 0.5(1 + \frac{s - \theta_t}{1 - \theta_t}) \quad (2)$$

이때, x, x', b 는 각각 추적된 박스, 검출기를 통해 재조정된 박스, 그리고 최종 객체 박스의 네 꼭짓점 좌표이다. x_{score} 는 추적된 박스에 대한 검출 스코어이다. 만약 검출 스코어가 문턱 값을 넘지 못하는 경우는 객체가 사라졌다고 보고 제거한다.

2.5 매칭되지 않은 검출된 객체 박스의 처리

매칭되지 않은 검출 결과는 잘못된 검출 또는 이전 프레임에는 없던 새로운 객체의 등장 두 가지 경우로 생각해 볼 수 있다. 전자의 경우 이후 프레임에서 추적을 수행하여 큰 성능 저하를 야기할 수 있다. 따라서 잘못된 검출의 영향을 줄이기 위해, 제안하는 기법은 보수적인 문턱 값 $\theta_d = 0.9$ 을 적용하여 검출 스코어가 문턱 값 이상인 경우에만 새로운 객체로 판단한다.

2.6 객체 박스의 중복 검사

마지막으로 동일 객체에 대한 중복된 박스를 제거하여 최종적인 객체 위치를 결정한다. 구체적으로 객체 박스들간의 IoU 비율을 계산하여 두 박스간 IoU 비율이 0.75 이상인 경우 중복으로 간주한다. 중복된 박스는 검출 스코어를 비교하여 높은 쪽 박스를 선택한다. 현재 프레임에서의 객체 위치를 결정하고 나면, [3] 논문의 방법을 통해 각 객체에

3. 실험 결과

본 논문에서는 정량적 평가를 위해 20개의 실험 영상에 대한 정확도 (Precision)와 재현율(Recall)을 측정하고 F1 score를 계산하여 평가지표로 사용하였다. 객체 검출 여부는 객체의 클래스를 올바르게 판단하고 검출 박스가 실제 객체의 박스와 IoU 0.5이상인 것으로 제한하였다.

표1의 결과는 고양이, 개, 말, 자동차, 사람 총 다섯 클래스에 대한 성능의 평균값이다. 기존 기법의 성능은 두 문턱 값(θ_d, θ_t) 이상의 검출 박스에 대한 성능을 각각 나타내었다. 제안하는 기법은 스코어가 θ_d 이상인 결과를 추적하기 시작하여 θ_t 이상인 결과들 중 선택하여 검출 결과로 사용한다. 기존 기법과 비교하여 θ_d 이상인 결과만 검출하였을 때보다 재현율이 높고 θ_t 이상인 결과를 모두 검출하였을 때보다 높은 정확도를 보였다. 또한 모든 경우와 비교하여 높은 F1 score 값을 보였다. 그림1에서 기존 기법과 제안하는 기법을 정성적으로 비교한다. 기존 기법의 경우 부정확한 객체 박스가 다수 출력되는 반면 제안하는 기법은 부정확한 객체 박스를 제거 및 재조정하여 정확한 결과를 반환하였다.

4. 결론

본 논문에서는 기존 객체 검출 기법[3]에 추적 기법[4]를 혼합하여 동영상에서 객체를 검출하는 시스템을 제안하였다. 또한 실험을 통해 제안하는 기법이 기존 기법에 비해 동영상에서의 객체 검출에 더 적합함을 보였다.

5. 참고문헌

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.
 [2] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In NIPS, 2015.
 [3] 문주혁, 김한울, 김창수, “다중 랜덤 워커를 이용한 객체 추적 기법”, 한국 방송·미디어 공학회 추계학술대회, 2016