

동영상을 위한 객체 검출 기법과 추적 기법의 결합

임경선, 김한울, 김창수

고려대학교

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

Joint Object Detection and Tracking in Video Sequences

Kyungsun Lim Han-Ul Kim Chang-Su Kim

Korea University

요약

본 논문에서는 동영상에서 제한된 종류의 동적 객체를 자동적으로 검출하여 추적하는 기법을 제안한다. 제안하는 기법은 객체 검출 기법[1]과 객체 추적 기법[2]의 협업을 통해 이를 수행한다. 검출기는 매 장면마다 객체들을 검출하고 이 중 높은 신뢰도의 객체에 대해 추적을 시작한다. 추적기는 이전 장면에서 학습된 분류기에 기반하여 객체를 추적한다. 추적 결과와 겹치는 검출 결과를 분석하여 현재 장면에서 객체의 정확한 위치와 모양을 추정한다. 겹치는 검출 결과가 없을 때는 검출기로부터 추적 결과의 신뢰도를 측정하고 문턱값에 따라 추적을 계속 진행하거나 종료한다. 실험 결과를 통해 제안하는 기법이 기존 검출 기법에 비해 우수한 검출 성능을 보임을 확인한다.

1. 서론

객체 검출 및 추적 기술은 영상처리 및 컴퓨터 비전 분야에서 중요한 연구 주제이다. 더욱이 최근에는 자동화 시스템에 대한 필요가 높아지고 있으며 영상 감시 시스템, 자동 주행 시스템 등에서 검출 및 추적 기술이 효과적인 영상 분석을 가능하게 한다. 객체 검출 기술과 추적 기술은 개별적으로도 개발이 되고 있으나, 본 논문에서는 각각 개발된 검출 기법과 추적 기법을 결합하여 서로 보완하도록 하는 알고리즘을 제안한다.

객체 검출 분야에서 컨볼루션 신경망(convolutional neural network)을 이용한 객체 검출 기법[1][3][4]들이 출현하면서 매우 큰 성능 향상이 있었다. 그러나 깊은 신경망을 사용하기 때문에 일반적으로 그 성능에 비하여 속도가 느려 동영상 적용에 어려움이 있다. 본 논문에서 사용한 Faster R-CNN[1]은 RPN(region proposal network)을 통해 객체 제안에 걸리는 시간을 줄여 실시간 적용이 가능하면서도 높은 성능을 보인다. Faster R-CNN은 영상에서 객체의 경계 박스(bounding box)를 높은 정확도로 추출하며 객체의 종류를 판별할 수 있다. 그러나 정지영상 분석을 위해 개발된 것이기 때문에 동영상에 적용하였을 때는 두 장면 사이에서 같은 객체를 연결시키지 못한다. 그리고 물체의 움직임으로 인한 변형에 민감하게 반응하여 같은 객체에 대해서도 장면마다 계산하는 스코어가 변동이 심하다. 이러한 불연속적인 결과들을 추적기를 통해 이어나갈 수 있다.

객체 추적 분야에서 배경으로부터 객체를 분리해내는 분류기 기반의 기법들은 전반적인 추적 성능 향상에 크게 기여하였다 [5]. 본 논문에서 사용한 SOWP[2] 추적 알고리즘은 객체 경계 박스 내에 존재하는 배경 및 객체 요소를 효과적으로 분리하여 높은 성능을 내는 추적기이다. 그러나 추적을 시작하기 위해서 초기 위치에 대한 외부 입력을 요구하며, 추적 종료 시점을 스스로 정할 수 없다. 또한 추적 중에 객체 박스의 크기를 조절할 수 없다.

두 기술의 장단점이 서로 보완되므로 이들을 조합한 시스템을 구성한다. 구체적으로 자동적으로 추적할 객체를 판별하여 추적을 시작할 수 있다. 객체의 모양이 변함에 따라 경계박스의 모양을 갱신하며 새로운 경계 박스를 이용하여 추적기의 분류기를 더욱 정확하게 학습할 수 있다. 객체가 사라짐을 판단하여 스스로 종료한다.

2. 제안하는 기법

2.1 추적 시작 및 진행

높은 정확도를 위하여 본 논문에서는 매 장면마다 Faster R-CNN이 검출한 박스 중에 이전 장면의 객체와 매칭이 되는 것을 제외한 나머지 중 스코어가 0.9 이상인 객체 박스에 대한 추적을 시작한다. 스코어 0.9 이상인 박스는 데이터셋에서 정확도(precision)가 0.919 이상으로 높은 값을 갖는 신뢰할 수 있는 박스들이다. 추적이 시작되면 SOWP는 첫 박스의 특성값을 업데이트하고 이에 기반하여 다음 장면에서의 객체 위치를 추적한다.

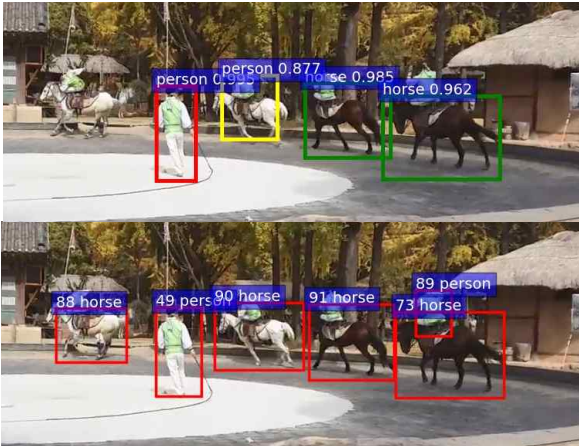


그림 1. 상: Faster R-CNN, 하: 제안된 기법. 상위 그림에서 검출 스코어 0.9이상인 결과는 빨강과 초록 박스 총 세 객체이다. 하위 그림에서는 이 외에도 세 객체를 더 검출하고 있다.

2.2 현재 장면 객체 위치 결정

이전 장면의 객체로 학습된 SOWP가 현재 장면에서의 객체 위치를 추적한다. 추적 결과는 위치적으로는 정확도가 높으나 객체 모양이 변형된 경우에 이를 인지하지 못한다. 따라서 추적 결과와 겹치는 검출 결과 박스를 통해 새로운 모양을 추정한다.

추적 결과와 겹치는 검출 결과란, 두 박스의 IoU(intersection over union) 비율이 0.5 이상이며 검출 스코어가 일정 값 이상인 검출 결과이다. 이때, 일정 값은 검출 결과에서 각 클래스 별로 영상 전체에 대해 정확도가 0.5 이상이 되는 스코어 threshold를 실험적으로 구한 값이다. 이러한 검출 결과가 있는 경우에는 객체가 현재 프레임에도 존재하는 것으로 판단하고 박스 모양을 재조정한다.

$$C_t = C_{t-1} + \sum_{k=1}^n f(S_t^k)(C_t^k - C_{t-1}) \quad (1)$$

여기서 C는 객체 박스의 좌표이고 S는 검출 스코어이다. f(x)는 검출 스코어가 일정 값 이상일 때는 0.7x를 쓴다. 이때 일정 값은 검출 결과에서 각 클래스 별로 영상 전체에 대해 정확도가 0.7 이상이 되는 스코어 threshold로 검출 결과 박스의 신뢰도를 높게 평가하여 위와 같은 함수를 적용한다. 일정 값 미만일 경우 신뢰도가 하락하기 때문에 0.7x²을 사용하여 효과를 줄인다.

추적 결과와 겹치는 검출 결과가 없는 경우에는 현재 프레임에 객체가 존재하는지에 대해 판단하기 위해서, Faster R-CNN의 객체 분류 부분인 Fast R-CNN[4]에 현재 장면과 추적결과를 입력하여 추적 결과에 대한 검출 스코어를 반환 받는다. 반환된 스코어가 위에서와 같은 일정 값 이상이면 추적 결과를 최종 객체의 위치로 결정하며, 그 외에는 객체가 더 이상 존재하지 않는 것으로 보고 추적을 종료한다.

2.3 객체 간 가림

여러 객체가 나타나는 영상에서는 객체들이 서로 겹쳐 하나가 다른 객체의 뒤로 사라지는 현상이 자주 일어난다. 두 결과 박스의 IoU가 0.5 이상일 때, 이러한 가림 현상이 일어나는 것으로 보고 두 박스 중 더 낮은 검출 스코어를 가지는 결과 박스를 삭제하고 추적을

	Precision	Recall	F1 score
Faster R-CNN	0.919	0.473	0.599
Proposed	0.883	0.657	0.723

표 1. Faster R-CNN과의 성능 비교
종료한다.

2.4 추적기 학습

최종적으로 현재 장면의 객체의 위치를 모두 결정한 후 추적기 SOWP의 분류기를 학습한다.

3. 실험 결과

본 논문에서는 정량적 평가를 위해 20개의 실험 영상에 대해 객체 검출 여부에 따른 정확도와 재현율(recall)을 측정하고 F1 score를 계산하였다. 객체 검출 여부는 객체의 종류를 올바르게 판단함과 동시에 제안 박스와 실제 박스의 IoU가 0.5이상인 것으로 한다.

제안하는 기법은 검출 스코어 0.9이상인 결과에서 추적을 시작하므로, 비교를 위해 Faster R-CNN 검출 기법에서도 검출 스코어 0.9이상인 것만 검출 결과로 사용하였다. 표 1 은 고양이, 개, 말, 자동차, 사람 총 다섯 개의 클래스에 대해 실험한 결과의 평균값이다. 제안된 기법은 Faster R-CNN과 비교하여 정확도가 약간 감소하였으나 재현율이 매우 상승하여 더 높은 F1 score를 얻었다. 비교 결과 더 정확하게 객체가 검출되었음을 확인할 수 있다. 그림 1 과 같이 제안된 기법에서는 기존 기법에 비해 더 많은 객체를 정확하게 찾아내고 있으며 순번을 부여하여 동영상에서 같은 객체들을 하나의 그룹으로 인식한다.

4. 결론

본 논문에서는 기존 객체 검출 기법[3]과 추적 기법[2]을 조합하여 동영상에서의 자동으로 객체를 검출하고 추적하며 스스로 종료 시점을 판단하는 시스템을 제안하였다. 실험을 통해 이 시스템이 검출 성능에 있어서 기존 검출 기법과 비교하여 우수함을 보였다.

5. 참고문헌

- [1] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In NIPS, 2015.
- [2] H.U. Kim, D.Y. Lee, J.Y. Sim, and C.S. Kim, "SOWP: Spatially Ordered and Weighted Patch Descriptor for Visual Tracking," In Proc. ICCV, 2015.
- [3] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.
- [4] R. Girshick. Fast R-CNN. In ICCV, 2015.
- [5] Y. Wu, J. Lim, and M.-H. Yang, "Object Tracking Benchmark," TPAMI, Sep. 2015.