Metric learning과 IoU 비교를 통한 객체추적 기법

*최인규 고민수 송 혁 **유지상
전자부품연구원, 광운대학교
*cig2982@keti.re.kr, kmsqwet@keti.re.kr, hsong@keti.re.kr, **jsyoo@kw.ac.kr

Object Tracking Technique with Metric Learning and IoU Comparison

*Choi, Inkyu Ko, Min-soo Song, Hyok **Yoo, Jisang
Korea Electronics Technology Institute, Kwangwoon University

요약

지속적인 딥러닝 기반의 영상처리 기술의 발전으로 객체분류나 객체검출 문제에 대해서 뛰어난 성능 보이고 있다. 하지만 객체추적 문제에서는 성능이 좋은 추적기는 실시간 동작이 불가능하고 딥러닝 기반의 객체추적도 단일 객체에만 고려한 기법이 많기 때문에 개선할 필요가 있다. 전처리로 검출된 객체영역과 kalman filter를 통해 예측된 추적영역 간의 embedding feature 비교를 통해 동일인물인지 판단하여 고유 ID를 부여하고 추적한다. 객체끼리 교차하거나 가려지는 상황에서 추적을 실패하게 되는데 이 후에 지속적인 추적을 위해 IoU 비교를 통해 후보 추적기로 남겨두는 과정을 거친다. 실험 결과 실시간 동작여부와 객체끼리 교차하거나 프레임 밖으로 나갔다가 다시 나타나는 경우에도 추적이 가능함을 확인하였다.

1. 서론

최근 딥러닝 기술의 발전으로 영상분석 및 이해의 성능이 향상됨에 따라 자율주행 자동차, 지능형 CCTV, 의료영상 분석 등 다양한 분야에 딥러닝 기반의 영상처리 기술이 적용되고 있다. 2012년 AlexNet[1]이 ImageNet Large Scale Visual Recognition Challenge(ILSVRC)를 우승한 이후로 지속적인 CNN 구조 및 알고리즘의 발전으로 영상분류(Image classification)는 top5 기준으로 96% 이상의 성능을 보인다. 그리고 객체검출(Object detection)도 Pascal VOC, COCO[2, 3]와 같은 데이터베이스에 대해 계속해서 객관적인 성능(mAP)이 향상되고 있다. 그 외에 얼굴검출, 행동인식, 영상분리 등 딥러닝 기반의 다양한 영상처리 기술들이 발전하고 있다. 그럼에도 불구하고 아직까지 보편화된 딥러닝 기반의 객체추적 기술은 없다. 보통 영상에서 단일 객체에 대해서만 다루거나 성능이 좋은 복수 객체 추적 기술의 경우 실시간 동작이 불가능하다. 인간의 안전을 위한 시스템에 적용할 때에는 실시간 추적기술이 불가피하기 때문에 처리속도가 매우 중요하다.

본 논문에서는 metric learning과 IoU(Intersection over Union) 비교를 통한 객체추적 기법에 대해 서술한다. Metric learning은 CNN을 통과한 embedding feature 간의 거리를 이용한 학습으로 동일 객체와 다른 객체 간의 분리 가능성을 높이는 방법이다. 영상 내 검출된 객체와 추적 영역간의 embedding feature를 비교하여 동일 인물 여부를 구별하여 추적한다. IoU는 검출 영역의 교차정도를 나타낸 것으로 객체가 교차하거나 가려졌을 때를 고려하여 추적 실패 시에 바로 해당 추적기를 버리는 것이 아니라 다시 사용할 수도 있는 후보 추적기로 두기 위해서 이용한다. 실험결과 복수 객체에 대해서 실시간으로 객체 추적이 가능하고 객체끼리 가려지거나 영상 프레임 밖으로 나갔다가 다시 들어온 경우에도 지속적으로 추적하는 것을 확인하였다.

2. 본론

객체추적을 위하여 영상 내 객체(사람)검출은 전처리로 진행되어야 한다. 본 논문에서는 YOLO v2[4]를 이용하여 객체를 검출하였다.

Metric learning을 위해 아래 그림 1과 같이 동일 인물에 대해 각각의 여러 사진을 같은 폴더에 묶어 데이터를 구성하도록 한다.

Figure 1. Configuring database for metric learning

여기서 사용할 방법은 triplet[5]으로 anchor, positive, negative의 세 가지 영상을 하나의 쌍으로 구성하여 CNN에 입력한다(anchor-positive 동일 인물, anchor-negative 다른 인물). 학습이 진행되면 그림 2와 같이 anchor와 positive 영상 간의 거리는 가까워지고 anchor와 negative 영상 간의 거리는 멀어지게 된다.

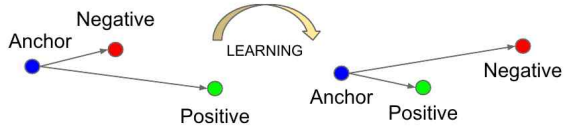

Figure 2. Learning process using Triplet

실제로 학습을 진행하면 아래의 그림과 같이 anchor-positive 간의 거리는 작아지고 anchor-negative 간의 거리는 멀어지게 된다. 그래서 테스트할 때 동일인물 구별을 위한 임계값을 anchor-positive 거리와 anchor-negative 거리 사이의 값으로 정하여 진행한다.

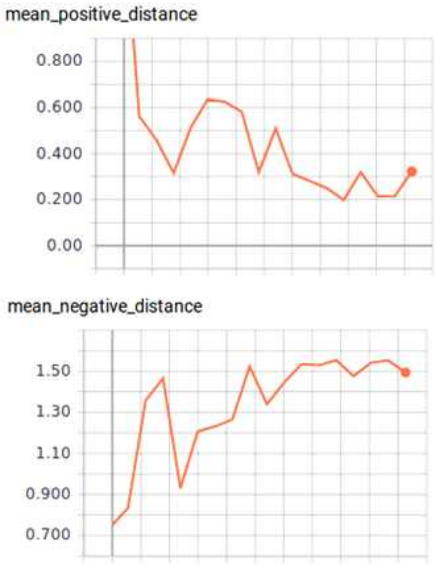

Figure 3. Learning results using Triplet

검출 영역과 kalman filter 기반의 예측 영역 간의 embedding feature를 비교하여 동일인물인 경우에 매칭 쌍을 구성한다. 객체 끼리 교차하거나 배경에 가려졌을 때 추적에 실패하는데 이후에 계속 추적하기 위해서 IoU를 비교하여 임계값보다 큰 경우에 후보 추적기로 둔다. 그림 4는 복수의 사람들이 나타나는 동영상에 객체 추적 기법을 적용한 결과이다. 객체가 교차할 때 ID가 교환되는 현상이 발생하기는 하나 교환 뒤에 바뀐 ID로 계속 추적하기 때문에 큰 문제는 아닌 것으로 판단하였다. 그리고 객체가 교차하거나 프레임 밖으로 나갔다가 다시 들어온 경우에도 기존의 ID로 계속하여 추적하는 것을 볼 수 있다.

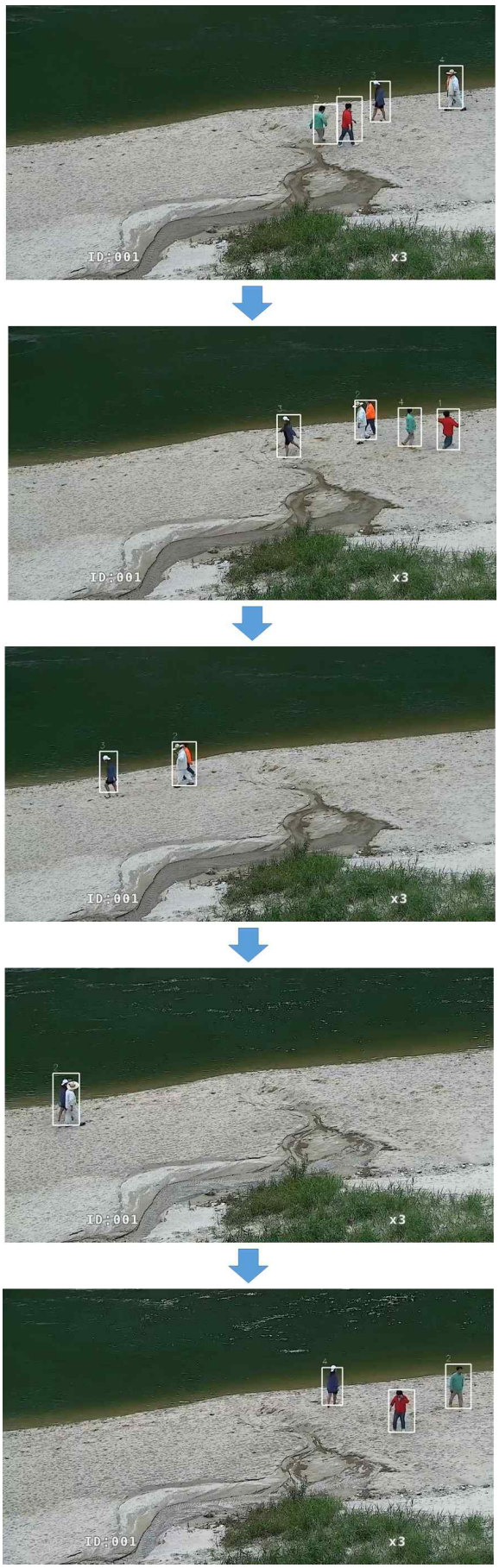

Figure 4. Multi-object tracking results in video

3. 결론

본 논문에서는 metric learning을 이용한 실시간 복수객체 추적 기법을 제안한다. Metric learning을 이용하면 CNN을 통과하여 추출된 embedding feature를 비교하여 동일인물과 다른 인물을 구분할 수 있다. 객체끼리 교차하거나 특정 물체에 가려지거나 할 때 추적에 실패할 경우가 생기는데 이러한 문제를 해결하기 위해 객체영역과 예측영역간의 IoU를 비교하여 다시 나타낸 객체에 대하여 재추적을 가능하도록 한다. 실험을 통하여 실시간 객체추적 여부와 객체가 가려지거나 프레임 밖으로 사라졌다가 등장하는 경우에도 추적이 가능함을 확인하였다.

ACKNOWLEDGMENT

본 논문은 2017년도 서울시 도시문제 해결형 기술개발 지원사업 (과제번호2016-시정-04) 의 지원을 받아 수행한 결과입니다.

참 고 문 헌

[1] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, 2012

[2] M. Everingham, L. Van Gool, C. K. Williams, J. inn and A. Zisserman, “The pascal visual object classes (voc) challenge”, International journal of computer vision, 88(2), 303-338, 2010

[3] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan and C. L. Zitnick, “Microsoft coco: Common objects in context”. In European conference on computer vision pp. 740-755, Springer, Cham, 2014, September

[4] J. Redmon and A. Farhadi, “YOLO9000: better, faster, stronger”, arXiv preprint, 2017

[5] F. Schroff, D. Kalenichenko and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.