

열영상에서 압축 도메인 정보를 이용한 딥러닝 기반 객체 탐지 방법

변주형, 남건욱, 박장수, 이종석, 심동규
광운대학교

{quswngud3, ngotic, jangsoopark, suk2080, dgsim}@kw.ac.kr

Deep-learning based Object Detection in Thermal Video Using Compressed-Domain Information

JooHyung Byeon, Gunook Nam, Jangsoo Park, Jongseok Lee, Donggyu Sim
Kwanwoon University

요 약

본 논문에서는 압축 영역에서 열 영상을 이용한 딥러닝 기반의 객체 검출 방법을 제안한다. 비디오 압축 표준인 High Efficiency Video Coding(HEVC)를 이용하여 부호화된 비트스트림으로부터 Intra Prediction Mode(IPM), Prediction Unit Size(PUS), Transform Unit Size(TUS)를 추출하고 3 채널 영상으로 변환하고 객체 검출 네트워크인 YOLO 에 입력으로 넣어주어 최종적으로 객체의 위치 및 객체의 종류를 예측한다. 실험 결과로써 복원된 열 영상과 검출된 결과를 주관적으로 보여줌으로써 압축영역에서 열영상을 이용한 객체 검출이 가능함을 보인다.

1. 서론

최근 Deep Neural Networks (DNN) [1], Convolutional Neural Networks (CNN) [2] 등과 같은 neural network 구조와 함께 구조에 맞는 학습법이 점점 고도화됨에 따라서 다양한 분야에서 딥 러닝이 사용되고 있다. 컴퓨터 비전(Computer Vision) 분야에서는 이를 이용하여 높은 성능을 내는 다양한 객체 검출 알고리즘들 상당한 성과를 내고 있으며 이러한 알고리즘들을 통한 많은 무인화 시스템들이 개발되고 있으며 상용화를 앞두고 있다.

일반적으로 객체 검출은 카메라를 통해 획득한 영상을 이용하여 객체 검출을 진행하고 기존 많은 연구들은 높은 정확도로 객체 검출을 수행하고 있다. 하지만, 앞서 말한 객체 검출 알고리즘들은 대부분 실내와 같은 고정된 조명 환경 또는 짧은 시간의 동일한 조명 조건에서 객체를 검출하고 있다. 따라서 조명 조건이 시시각각 변화하는 실외 또는 자동차에 부착된 카메라와 같은 환경에서는 동일한 객체 검출 알고리즘 사용이 어렵다. 조명에 강인한 객체 검출을 위해서 적외선 센서를 이용 획득한 열 영상을 이용해 객체 검출을 하는 응용 분야도 있다. 적외선 센서를 통해 영상을 촬영하여 물체가 자체적으로 발산하는 복사 에너지(온도)차이를 감지하여 영상으로 재현하는 열 영상은 조명 환경에 무관하게 양질의 데이터를 취득할 수 있다. 이러한 특성으로 인해 열 영상은 국방 감시, CCTV 등의 야간 감시에 많이 활용되고 있으며 자율주행 자동차의 가장 중요한 기술 중 하나인 나이트비전 [3]에도 필수적으로 사용된다. 열 영상의 취득은 대부분 CCTV 형태의

네트워크 카메라를 통해 이루어지고 있다. 네트워크 카메라의 동작은 영상을 취득 후 자체 DSP 칩을 통해 실시간 부호화 후 네트워크에 전송하고, 전송 후 복호화 하여 영상을 디스플레이 장치에 전달해 준다. 이와 같은 구조에서는 영상 복호화 과정이 필수적으로 요구되며 이로 인한 지연 현상이 발생하게 된다. 열 영상 사용의 주된 목적은 실시간 감시인데 실시간 처리에서의 지연 현상은 치명적으로 작용하게 된다. 지연 현상을 해결하기 위해 bitstream 을 이용하여 바로 객체를 검출하기 위한 연구들이 진행되고 있다 [4]. 하지만 열 영상을 이용한 연구는 아직 전무한 실정이다. 따라서 본 연구에서는 부호화된 열 영상 bitstream 에서 Intra Prediction Mode (IPM), Prediction Unit Size (PUS), Transform Unit Size (TUS)를 부분 복호화 하여 이를 이용한 딥러닝 기반 객체 검출 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2 절에서는 제안하는 방법을 설명하고 3 절에서는 제안한 방법의 정확도를 실험을 통해서 확인한다. 마지막으로 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 제안하는 방법

본 논문에서 제안하는 열 영상을 이용한 객체 검출 방법은 부호화된 압축 도메인의 비트스트림을 디코딩하면서 생성되는 부가 정보들 중 화면 내 예측 방식으로 부호화된 프레임의 PUS, TUS, IPM 으로 구성된 이미지로 변환하여 총 3 채널의 영상을 생성하고 실시간 객체 검출에 사용되는 딥 러닝 네트워크인 YOLOv3 의 입력으로 넣어준다. 객체 검출의

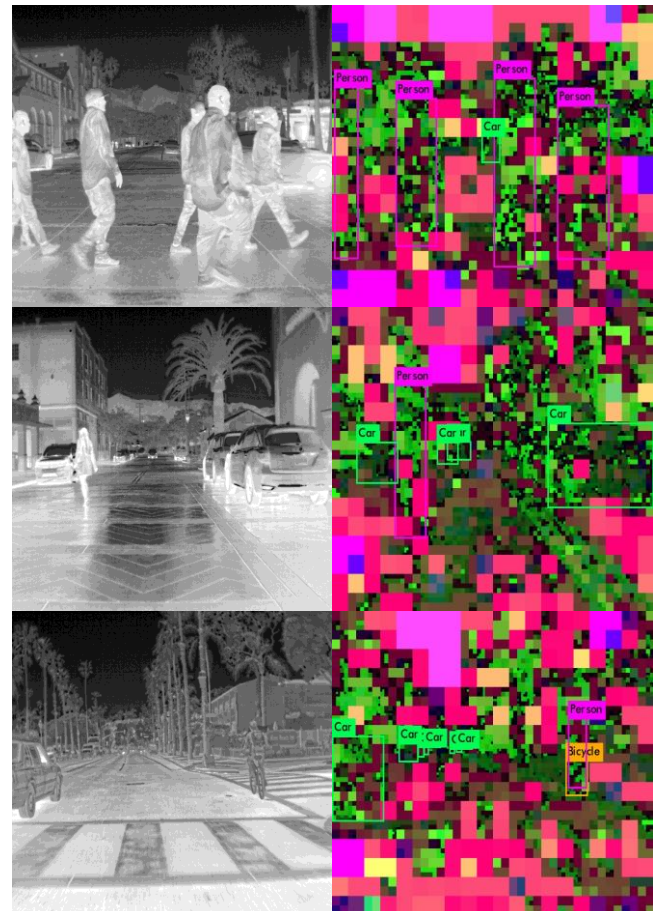
결과는 YOLOv3 의 출력 텐서(tensor)를 후 처리 과정인 non-maxima suppression 을 통해서 바운딩 박스(bounding box)와 클래스(class)를 예측한다. 그림 1.은 제안하는 방법의 흐름도이다. 비디오 압축 표준인 HEVC 로 부호화된 비트스트림을 CABAC(Context Adaptive Binary Arithmetic Coding) 복호화 하여 나온 IPM, PUS, TUS 를 추출하여 영상으로 변환한다. 영상은 일반적으로 0~255 의 값을 지니기 때문에 비트스트림에서 추출한 각각의 정보를 매핑해주어야 한다. HEVC 는 총 35 개의 화면내 예측 모드를 사용하고 그 값은 0~34 의 값을 갖는다. 따라서, 0~34 의 값을 0~255 로 값을 매핑해 사용하며, 화면내 예측 모드는 PU 단위로 추출되기 때문에 영상 생성 시 PU 내부에 하나의 화면내 예측 모드로 채운다. 그리고 PU 크기는 4, 8, 16, 32, 64 를 갖을 수 있고 각 크기는 0, 17, 51, 119, 255 로 맵핑 하고 4, 8, 16, 32 의 크기를 갖는 TU 의 경우 0, 36, 109, 255 로 맵핑 하여 영상 생성한다. PU, TU 크기도 화면내 예측과 같이 단 단위 내부에 동일한 값으로 채워 넣어 영상을 생성한다.

2. 실험 결과

본 논문에서는 제안하는 방법의 성능을 평가하기 위하여 HEVC 의 레퍼런스 소프트웨어인 HM-16.19 부호화기를 이용하여 열 영상을 부호화 하여 비트스트림을 생성한다. 네트워크 학습 과정에 사용하는 열 영상은 FLIR 사에서 제공하는 FLIR Thermal Starter Dataset 을 이용한다. 이 데이터는 16 비트 영상이기 때문에 부호화를 위하여 8bit 로 변환하여 HEVC 로 부호화 한다. 이후 부호화된 비트스트림에서 IPM, PUS, TUS 를 추출하기 위해 HM-16.19 복호화기를 수정하여 IPM, PUS, TUS 추출기 및 영상 변환기를 구현하였다. 최종적으로 앞의 과정으로 생성된 영상을 통해 YOLO 네트워크를 학습한다. 학습에 사용된 영상은 총 4000 장이고 'Person', 'Car', 'Bicycle'로 분류되어 있다. 그림 2 는 본 논문에서 제안한 객체 검출 방법의 검출 결과이다. 그림 2(a)는 열 영상의 비트스트림을 복원한 결과이고 그림 2(b)는 비트스트림을 이용하여 객체 검출한 결과이다.

3. 결론

본 논문에서는 열영상을 이용한 객체 검출에서 지연 현상 감소를 위하여 압축 영역에서 열영상을 이용한 딥러닝 기반 객체 검출 방법을 제안하였다. 본 논문에서 제안한 방법은 HEVC 로 부호화된 열영상의 비트스트림으로부터 IPM, PUS, TUS 를 추출하여 3 채널 영상으로 변환하고 학습된 YOLO 를 이용해 객체 검출한다. 제안한 방법은 주관적 객체 검출 결과를



(a) Reconstructed IR image (b) Detection results
 그림 2. 제안하는 방법의 객체 검출 결과

통해 압축 영역에서 열 영상을 이용한 객체 검출이 가능함을 보였다.

감사의 글

본 연구는 2018 년도 광운대학교 대학중점연구소 지원 사업 및 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT 연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2018-2016-0-00288)

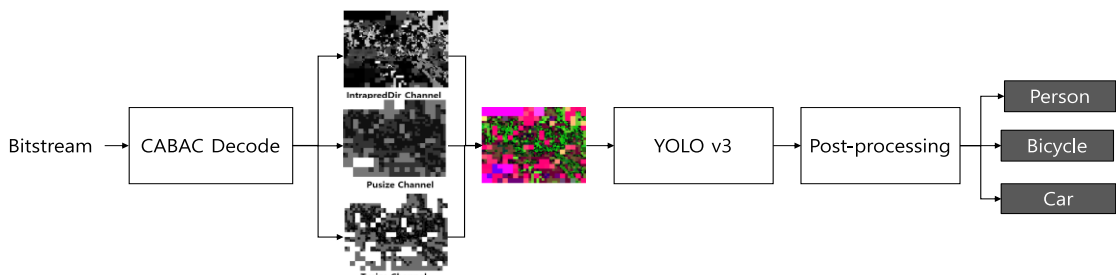


그림 1. 제안하는 열영상에서 딥러닝 기반 객체 검출 방법의 흐름도

참고문헌

- [1] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, “Scalable object detection using deep neural networks,” CVPR, 2014.
- [2] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet classification with deep convolutional neural networks,” NIPS, 2012.
- [3] F. Xu and K. Fujimura, “Pedestrian Detection and Tracking with Night Vision,” Proc. IEEE Intelligent Vehicle Symp., 2002.
- [4] S. R. Alvar, H. Choi, and I. V. Bajic, “Can you find a face in a HEVC bitstream?,” IEEE ICASSP’18, Apr. 2018.
- [5] FLIR System(2018). FREE FLIR Thermal Dataset for Algorithm Training. Retrieved from <https://www.flir.com/oem/adas/adas-dataset-form/>