

전 방향 객체 검출을 위한 360도 파노라마 영상에서의 CNN기반 객체 검출 방법

신동훈, 윤국진
 광주과학기술원 전기전자컴퓨터공학부
 {soogood114, kjyoon}@gist.ac.kr

요약

본 논문에서는 360도 파노라마 영상에서 전 방향 객체에 대한 CNN기반의 객체 검출 방법을 제시한다. 또한, 파노라마 영상 내의 객체들 가운데 좌우 끝 경계에 의해 분할된 객체에 대한 대처방법에 대해 기술한다. 제안한 방법에서는 기존의 일반 카메라 영상과 비교해 객체의 왜곡이 적은 360도 파노라마 영상을 대상으로 객체 검출을 수행한다. 하지만, 360도 파노라마 영상은 영상의 끝에 위치한 객체가 좌우 최외각에 의해 분리가 되는 문제가 생길 수 있다. 이에 본 논문의 방법은 CNN 기반의 객체 검출 기법 분석을 통해, 파노라마 영상 좌우 끝에 위치하는 분할된 객체 크기의 최대치를 추정한다. 추정된 최대 객체 크기만큼 원 영상을 확장함으로써 파노라마 영상 좌우 끝에 분할된 객체를 합쳐진 하나의 객체로 복원한다. 마지막으로, 확장된 영상에서 객체 추정을 한 다음, 원 영상 크기에 맞게 그 추정된 결과를 후처리하여 결과적으로 원 영상에 대한 전 방향 객체 추정을 수행할 수 있다.

1. 서론

최근 가상현실(VR) 시장이 성장하면서 360도 카메라의 시장이 커지게 됨에 따라 이를 활용한 영상 처리 및 컴퓨터 비전 분야의 연구가 활발히 진행되어지고 있다. 특히, 4차 산업 혁명 분야 중 하나인 첨단 운전자 보조 시스템(Advanced Driver Assistance Systems, ADAS)에서도 전 방향 상황인지를 위해 360도 카메라를 채용하는 경우도 많아지고 있어 이에 맞는 컴퓨터 비전 기술 적용이 필요한 실정이다.

본 논문에서는 영상 내의 객체의 식별, 위치정보, 상황인식 등에 필수적인 객체 검출 기법을 360도 카메라 영상에 적용하는 방법을 제안한다. 제안된 방법은 일반 카메라 영상에 맞는 기존에 있는 최고 성능의 객체 검출 방법을 응용하고 추가적 후처리를 통해, 파노라마로 표현된 360도 영상을 대상으로 전 방향 객체검출을 수행한다.

2. CNN 기반 객체 검출 기법

최근 CNN을 통해 객체 검출 성능과 속도를 획기적으로 향상시킬 수 있었다. 본 논문에서는 이러한 CNN을 활용한 객체 검출 기법 중 SSD: Single Shot MultiBox Detector[1]를 객체 검출 기법으로 사

용하여 다양한 스케일의 객체에 대해 빠르고 정확한 검출이 가능하도록 했다. 그리고 SSD내에 완전 연결 계층(fully connected layer)이 없기 때문에 입력 영상의 크기에 대한 제한이 없다. 이러한 특징들은 다른 영상에 비해 크기가 큰 360도 파노라마 영상에서의 객체 검출을 성공적으로 할 수 있게끔 한다.

본 방법에선 SSD-512모델을 객체 검출기로 사용한다. 더 다양한 스케일의 객체들을 더 정확하게 검출할 수 있게 하였다. 해당 SSD의 학습은 다양한 객체 종류가 있는 POSCAL VOC 데이터 셋과 COCO 데이터 셋으로 진행하여 범용성이 있는 객체 검출 수행이 가능하도록 했다.

3. 온전한 객체 검출을 위한 영상 확장

파노라마 영상은 일반 카메라 영상과 비교하여 크게 다르지 않아 기존의 성능 좋은 객체 검출기를 활용할 수 있지만, 객체가 파노라마 최외각에 걸쳐 존재하게 되는 경우 객체가 좌우로 나뉘어 보이게 되는 문제점이 있다. 이는 기존의 객체 검출기로 검출하게 되면 하나의 객체임에도 불구하고 1 또는 2개의 부정확한 객체 검출, 아니면 아예 검출이 안 되는 경우도 생기게 되어 온전하게 전

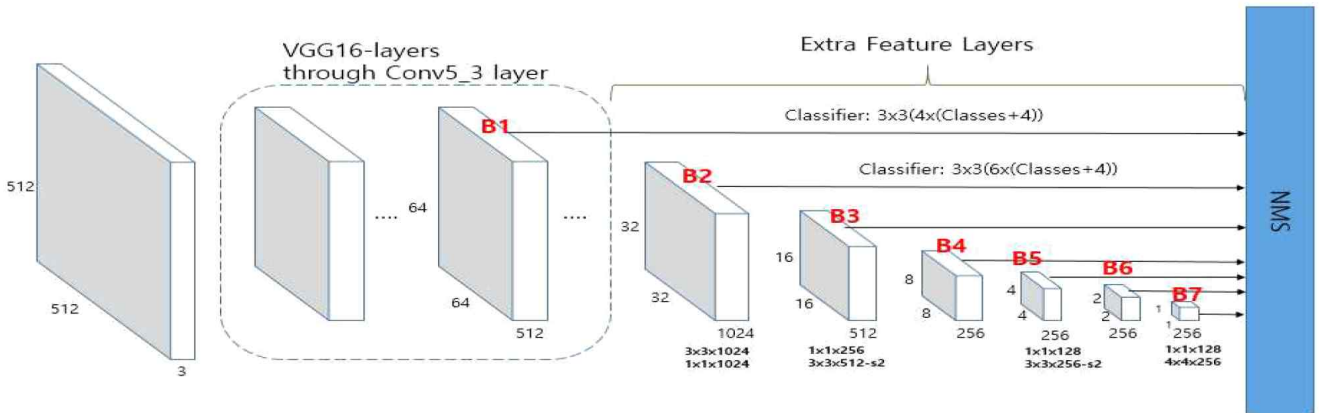


그림 1. SSD-512 구조

방향의 객체 검출이 수행될 수 없다. 그림3~5의 위쪽 그림들은 앞서 언급한 파노라마 영상에 의해 생길 수 있는 문제들의 예시이다. 이에 제안된 방법은 SSD의 최대 수용 영역(receptive field)[2]을 참고하여 영상 좌우 최외각에 위치하는 분할된 객체 크기의 최대치를 추정하여 그 최대치 크기의 넓이만큼 원 영상을 우측으로 확장하여 어떤 분리된 객체든지 완벽히 복원하고자 하였다.

본 논문에 제안된 방법에서는 분할된 객체의 최대 크기는 분할되지 않은 객체의 크기이기 때문에 가장 큰 객체 추정이 이루어지는 영역의 크기를 분할된 객체 크기의 최대치로 여겼다. 이러한 최대치를 구하기 위해 SSD의 feature map들 중 객체 추정이 이루어지는 특정 feature map들을 살펴볼 필요가 있다. 그림1에서 이러한 특정 feature map들을 B1~7으로 표시 했으며 본 논문에서는 통칭으로 key feature map이라 명명하겠다. 그림 1과 같이 SSD는 객체의 특정한 스케일 정보를 가지고 있는 여러 개의 key feature map의 픽셀단위로 객체 추정이 이뤄진다. 그러므로 픽셀 단위의 수용 영역이 가장 큰 key feature map을 찾아야 한다. 그림 1을 보면 feature map이 SSD의 네트워크 안에 있는 여러 번의 convolution과 max pooling 연산을 통해 영상의 정보가 함축 되어가면서 그 크기가 종단으로 가면 갈수록 더 작아지는 것을 확인할 수 있다. 이는 그 만큼 많은 양의 영상 정보가 한 픽셀 당 함축이 되어 있다는 것을 의미한다. 따라서 가장 크기가 작은 key feature map의 픽셀에서 가장 큰 수용 영역을 갖고 있으며 이 수용 영역 내에 크기가 가장 큰 객체들의 추정이 이루어진다.

표 1. SSD-512의 key feature map 크기 변화

계층	스케일 변화
원 영상($M \times N$)	-
B1 ($M_1 \times N_1$)	$\lceil M/8 \rceil \times \lceil N/8 \rceil$
B2 ($M_2 \times N_2$)	$\lceil M_1/2 \rceil \times \lceil N_1/2 \rceil$
B3 ($M_3 \times N_3$)	$\lceil M_2/2 \rceil \times \lceil N_2/2 \rceil$
B4 ($M_4 \times N_4$)	$\lceil M_3/2 \rceil \times \lceil N_3/2 \rceil$
B5 ($M_5 \times N_5$)	$\lceil M_4/2 \rceil \times \lceil N_4/2 \rceil$
B6 ($M_6 \times N_6$)	$\lceil M_5/2 \rceil \times \lceil N_5/2 \rceil$
B7 ($M_7 \times N_7$)	$(M_6-1) \times (N_6-1)$

표 1은 SSD-512모델의 key feature map(B1~7)의 단계별 크기 변화를 나타낸다. 이를 통해, 가장 작은 B7의 크기를 쉽게 구할 수 있고 이 크기와 원 영상의 크기를 비교하여 넓이 스케일 변화량과 높이 스케일 변화량을 구할 수 있다. 예를 들어, 원 영상의 크기가 1024x1024이고 B7에 해당하는 feature map의 크기가 3x3라고 하면 높이 스케일 변화, 넓이 스케일 변화 모두 약 341이 된다. 이렇게 구한 넓이 스케일 변화량과 높이 스케일 변화량은 각각 최대 수용 영역의 넓이와 높이가 된다. 즉, 간단한 연산을 통해 최대 수용 영역을 구함으로써 구하고자 하는 분할된 객체 크기의 최대치를 얻어낼 수 있다.

만약 원 영상과 B7사이의 넓이 스케일 변화량을 D라고 한다면, 그림 2처럼 원 영상에서 x값이 0에서부터 D사이에 해당하는 영상을 복사하여 얻어진 영상을 본 영상 우측 최외각에 붙여 넣는 방법으로 원 영상을 확장하는 작업을 한다. 이를 통해, 어떤 크기의 객체가 영상의 좌우 최외각에 의

해 분리되었든 간에 온전하게 그 객체가 온전하게 복원되도록 하였다.

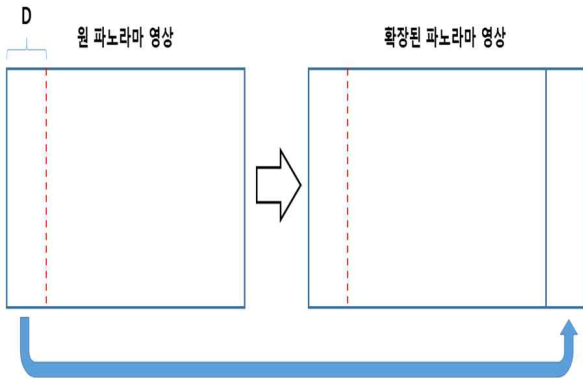


그림 2. 파노라마 영상 확장

4. 360도 파노라마 영상에서의 객체 검출

SSD는 실시간 객체 추정을 위해 입력 영상의 크기에 대한 제한이 있다. 예를 들어, SSD-512를 사용한다면 어떤 입력 영상이든지 영상의 크기가 512x512 크기로 와핑(warping)된 다음, 객체 추정이 이루어진다. 하지만, 상대적으로 크기가 큰 360도 파노라마 영상에서의 이러한 제한은 많은 영상 정보 손실을 가져오기 때문에 정확한 객체 추정이 이루어 질 수 없다.

따라서 본 논문의 방법에서는 앞서 명시한 기법대로 입력 영상 확장한 다음, 추가적 영상 와핑기법 없이 확장된 360도 파노라마 영상 크기에 맞게 SSD기반 객체 검출을 수행하도록 하였다. 이를 통해, 360도 파노라마 영상을 대상으로 더 정확한 객체 추정이 가능하게끔 했다. 그리고 객체 추정 결과는 객체 영역의 외각선의 x, y축의 최소, 최대값을 파라미터로 바운딩 박스(bounding box)형태로 나오도록 하였다. 즉, 객체가 검출이 되면 바운딩 박스는 (x_{min}, y_{min}, x_{max}, y_{max})형태로 나오고 해당 객체에 대한 검출 신뢰 값(confidence)도 함께 나온다.

5. 전 방향 객체 검출 후처리

얻어진 객체추정 결과물은 원 영상이 아닌 확장된 영상의 결과이므로 원 영상에 해당하는 검출결과들만 추출하는 후처리 과정이 필요하다. 원 영상이 360도 파노라마이기 때문에 본 방법에선 원 영

상의 우측 최외각 x좌표와 같은 확장된 영상의 x좌표를 w라고 하고 검출된 객체의 바운딩 박스의 x_{max}값이 w보다 작은 객체 검출들만 추출한다. 만약 추출되지 못한 바운딩 박스들 중 x_{min}과 x_{max} 사이에 w가 있는 박스가 있다면 이는 우리가 처리하고자한 영상 좌우 최외각에 의해 분리된 객체에 대한 검출 결과이므로 추가적인 처리가 필요하다.

이를 제안된 방법에서는 원 영상이 해당하는 바운딩 박스 좌표 값 중 x_{max}를 |x_{max} - w|로 교체하도록 하여 원 영상의 왼쪽 공간에 위치시켜 원 영상의 객체를 가리키도록 하고, 온전한 객체에 대한 검출이므로 객체 검출 신뢰 값은 그대로 유지시키도록 했다. 원 영상이 360도 파노라마 영상이기 때문에 이러한 방식으로 분리된 객체 검출 결과를 온전히 합쳐진 객체에 대한 검출 결과와 검출 신뢰 값으로 대체가능하다. 마지막으로, 분할된 객체에 대한 잘못된 검출 결과들은 위 과정을 통해 얻어진 온전한 검출 결과들과 겹쳐지기 때문에 최종 검출 결과에 제외시키도록 하였다.

6. 실험 결과 및 분석

360도 객체 검출 결과를 확인하고 분석하기 위해 가상 환경에서 얻은 4 방향 영상들이 포함된 Synthia dataset[3]을 이용하였다. 이 데이터베이스는 비록 사이버 가상 환경에서 얻어진 영상들이지만 다양한 환경과 다양한 객체들이 포함된 영상들을 다수 보유하고 있다는 장점이 있다. 이 데이터베이스로부터 얻은 4 방향 영상을 equi-rectangular image로 360도 파노라마 영상을 만들고 제안한 방법에 적용 후 그 결과를 확인하였다.

그림 3~5는 분리된 객체에 의해 생길 수 있는 3가지 문제를 객체 검출 결과 비교를 통해 나타낸 것이다. 결과 비교는 두 개의 영상 결과를 가지고 이루어 졌으며, 위 쪽 영상은 전 방향 객체를 검출 하되 파노라마 영상 외각에 분리된 객체에 대한 대처가 전혀 안 이루어진 결과 영상이고, 아래쪽 영상은 앞서 설명한 방법을 통해 분리된 객체에 대한 대처가 이루어져, 분리된 객체가 하나의 객체로 검출이 된 결과 영상이다.

그림 3의 경우는 분리된 객체가 어느 쪽에서도 검출이 안 되는 경우이다. 보통 작은 객체가 분리된 경우 이와 같은 문제가 발생한다. 그리고 그림 4의 경우는 객체 검출이 하나로 나오고 그 객체의 검출 신뢰 값도 어느 정도 나오지만 잘린 영역

들 중 한쪽 영역에만 검출이 되었기 때문에 객체의 위치 정보가 부정확하게 나오는 문제점이 있다. 이런 경우는 객체가 심하게 비대칭적으로 분리가 될 때 주로 발생한다. 마지막으로, 그림 5의 경우가 가장 빈번한 문제이다. 하나로 나와야 하는 객체 검출이 2개로 나오고 그 2개의 위치 정보마저 부정확하기 때문에 하나의 정확한 객체정보만 나오도록 해야 한다.

위에서 열거한 문제의 경우 모두 제안한 방법대로 영상 확장을 통해 분리된 객체를 하나의 객체로 복원을 시켜 객체 검출을 수행하게 되면, 하나의 온전한 객체 결과로 나올 뿐만 아니라 그 객체의 더 정확한 위치 정보와 더 높은 검출 신뢰 값도 얻을 수 있게 된다.

7. 결론

본 논문은 360도 파노라마 영상을 대상으로 SSD 객체 검출기를 응용 및 적용하여 전 방향 객체 검출을 할 수 있는 방법을 제안하였다. 특히, 본 논문에서는 파노라마 영상의 특성상 영상의 좌우 최외각 선에 의해 하나의 객체가 2개의 객체로 나뉘진다면 객체 검출에 문제가 생길 수 있다는 점을 주목하였다. 이러한 문제를 해결하기 위해, 원 영상을 객체가 생길 수 있는 최대 크기만큼 원 파노라마 영상을 좌측에서 복사하여 우측으로 확장하고 그

확장된 영상을 대상으로 SSD 객체 검출기를 통해 전 방향 객체들을 온전하게 검출할 수 있는 방법을 제안하였다. 확장된 영상으로부터 검출된 객체 정보를 후처리하여 원 파노라마 영상에 맞는 최종적 객체 추정 결과를 도출하고 그 결과를 바운딩 박스와 검출 신뢰 값 형태로 나오도록 하였다. 확인해 본 결과, 본 방법에 의해 하나의 객체가 영상의 최외각에 의해 2개로 검출되는 문제가 사라졌으며 그 객체에 대한 검출 신뢰 값 또한 대체로 향상시킬 수 있었다.

참고문헌

- [1] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg, "SSD: Single Shot MultiBox Detector", ECCV, 2016.
- [2] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel "Understanding the Effective Receptive Field in Deep Convolutional Neural Networks"
- [3] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M. Lopez, "The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes", CVPR, 2016

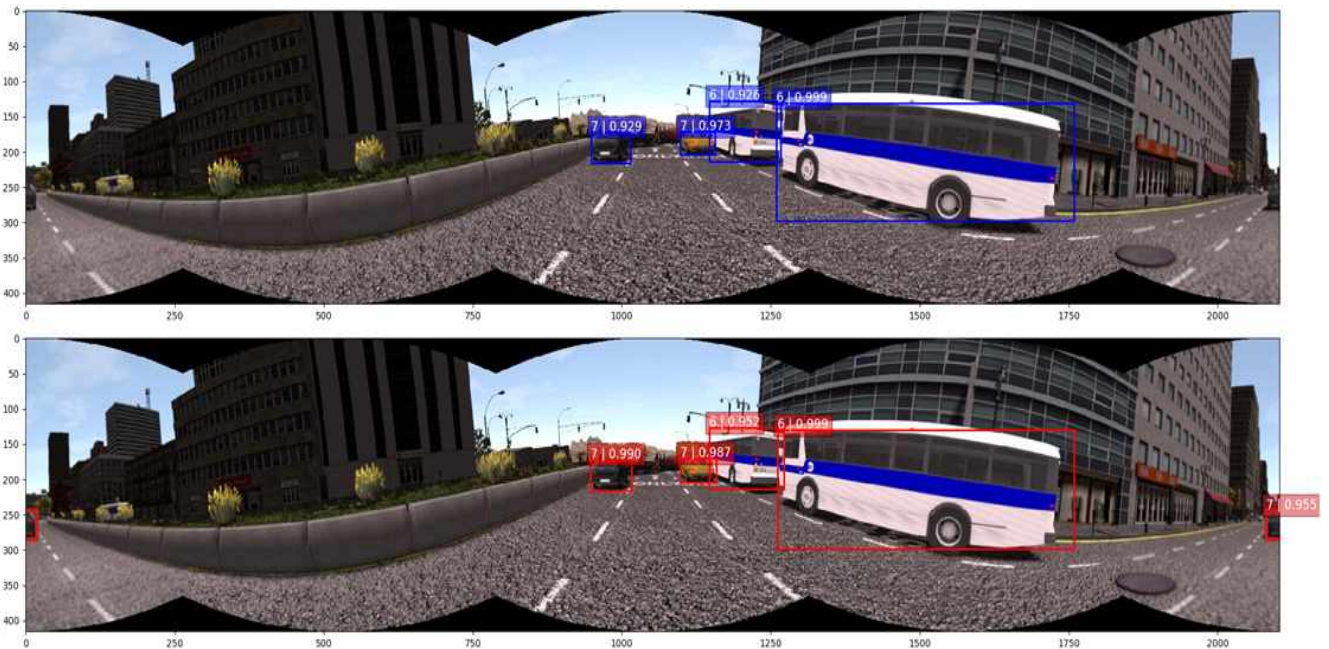


그림 3. 분리된 객체에 대한 검출을 실패한 경우(상), 분리된 객체에 대한 대처가 된 경우(하)

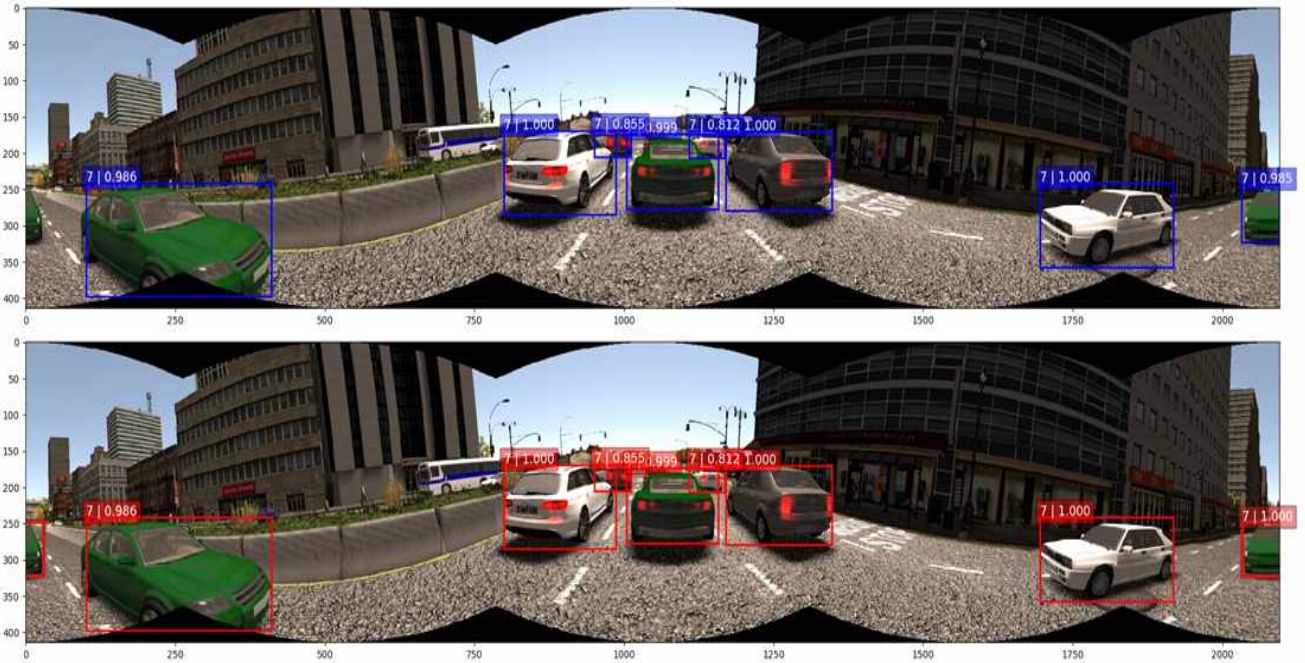


그림 4. 분리된 객체에 대한 부정확한 검출이 된 경우(상), 분리된 객체에 대한 대처가 된 경우(하)

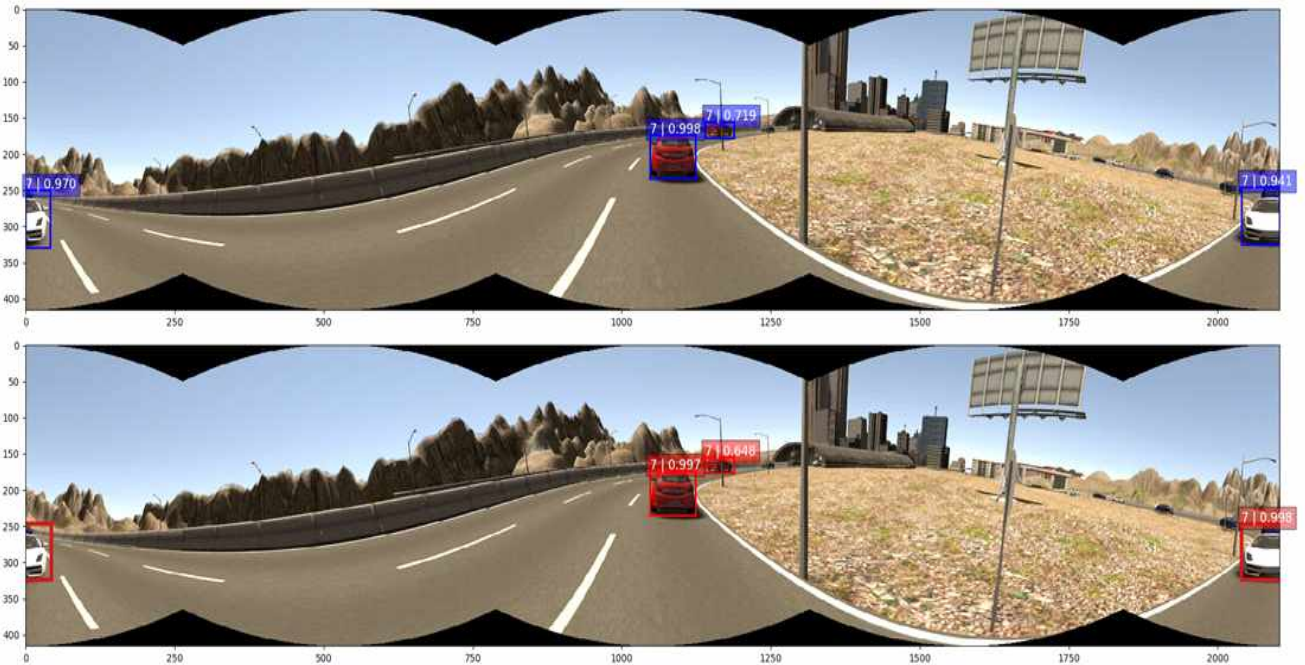


그림 5. 분리된 객체에 대한 중복 검출이 된 경우(상), 분리된 객체에 대한 대처가 된 경우(하)