

# 터널 CCTV 영상의 심한 원근현상 극복을 위한 역 원근변환 기법의 객체인식 성능에 미치는 영향 검토

이규범\*, 신휴성\*\*

\*한국과학기술연합대학원대학교, 스마트도시건설융합과

\*\*한국건설기술연구원, 미래스마트건설연구본부

e-mail:kyubeomlee@kict.re.kr, hyushin@kict.re.kr

## A Study on the Influence of Inverse Perspective Transformation Technique to Overcome Severe Perspective of Tunnel CCTV Images on Object Detection Performance

Kyu Beom Lee\*, Hyu Soung Shin\*\*

\*Smart City Construction Convergence, University of Science and Technology

\*\*Dept. of Future & Smart Construction Research, Korea Institute of Civil Engineering and Building Technology

### 요약

터널 내 설치된 CCTV는 터널 구조물의 공간적인 한계로 인해 낮은 높이로 설치되어 있으며, 150m 이상 간격으로 설치되어 있기 때문에 원근현상이 심하다. 이 CCTV를 기반으로 하여 터널 내 사고를 탐지할 수 있는 터널 영상유고 시스템이 운용 중에 있는데, 기존 터널 영상유고 시스템은 원근현상으로 인해 터널 CCTV로부터 멀리 떨어진 차량, 보행자 및 화재와 같은 객체를 검출하기 어렵다. 이 문제를 해결하기 위해 관심영역 설정 및 역 원근변환(inverse perspective transform) 기법을 도입하였으며, 이 과정을 통해 얻은 변환영상은 먼 거리에 있는 객체가 확대됨으로 인해 터널 영상유고 시스템이 검출하는 것이 가능하다. 이 과정이 유효하다는 것을 증명하기 위해, 터널 CCTV 영상의 원본영상과 변환영상을 바탕으로, 각각 동일한 조건을 가지는 데이터셋 제작, 딥러닝 객체인식 모델의 학습 및 추론을 통해 객체인식 성능을 비교하였으며, 결과적으로 변환영상을 사용한 경우가 먼 거리에서 차량에 대한 객체인식 성능이 뛰어난 것을 보였다.

CCTV와 멀어질수록 영상에서 보이는 물체의 크기가 매우 작아지므로 영상유고 시스템의 성능이 하락한다는 문제점이 있다.

## 1. 서론

도로 터널에서 발생하는 사고는 대피공간의 부족, 어두운 환경으로 인해 2차 사고의 발생위험이 크다. 그렇기 때문에 사고의 발생을 신속하게 인지한 다음 대처할 수 있어야 하며, 이를 보조하기 위해 200m 이상의 터널 연장을 가진 터널 현장은 CCTV의 설치 및 감시를 통해 사고의 탐지 및 대처하도록 되어 있다. 일부 터널 현장에서는 터널 내 CCTV를 기반으로 정차, 역주행, 보행자 및 화재 유고상황을 탐지할 수 있는 영상유고 시스템을 운용중이다.

터널 내 CCTV의 설치간격은 국가마다 세부적으로 다르지만 최소 75m, 보통 150m~200m의 간격으로 설치되며, 터널 CCTV 영상의 가시거리도 설치간격에 맞추어 각도 및 초점을 설정한다. 그런데, 이러한 방법으로 설정하게 되면 터널 CCTV의 원근현상이 심해지며, 터널 CCTV의 설치높이가 구조물의 한계로 인해 낮기 때문에 물체의 실제 위치가 터널

## 2. 적용된 이론

### 2.1 역 원근변환 기법 (Inverse perspective transform)

이 문제를 해결하기 위해 본 논문에서는 역 원근변환 기법을 도입하였다. 역 원근변환 기법은 영상 변환의 종류 중 하나로, 주어진 영상에서 심한 원근현상을 보정하여 완화하는 영상 변환 방법을 말한다. 현재 자율주행 분야에서 사용되고 있으며, 본 논문에서는 터널 CCTV 영상의 심한 원근현상을 완화하기 위해 이 방법을 사용하였다.

먼저, 역 원근변환 기법을 적용하기에 앞서 원본 터널 CCTV 영상 (이하 원본영상)에서 관심영역을 설정하였으며, 다음 절차와 같이 진행하였다.

(1) 먼저 원본영상에서 차량 주행방향과 평행하고, 차선의

맨 바깥쪽 직선 두 개를 그린다.

(2) 터널 CCTV에서 가장 가까우며, 실제로 차량 주행방향과 수직인 직선과 (1)에서 구한 두 직선 사이 교점 2개를 구한다.

(3) 다른 CCTV와 설치간격만큼 멀리 떨어져 있으면서 실제로 차량 주행방향과 수직인 직선과 (1)에서 구한 두 직선 사이 교점 2개를 구한다.

이러한 절차를 통해 관심영역 및 4개의 점을 설정할 수 있으며, 역 원근변환을 실행하기 위해 관심영역 4점과 관심영역 4점에서 수평 및 수직의 최소 및 최대점으로 이루어진 직사각형 4 꼭지점간의 위치변환 관계를 표현하는 호모그래피(homography) 행렬이 필요하며, 다음 식 (1)과 같이 표현할 수 있다.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \quad \text{식 (1)}$$

식 (1)에서 H는 호모그래피 행렬을 말한다. 그리고 원본영상 좌표계의 점 (x,y)가 있고, 변환영상 좌표계의 점 (x',y')이 있을 때, 다음 식 (2)와 같이 역 원근변환을 통해 원본영상에서 변환영상으로 좌표계 변환이 가능하다.

$$S \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \text{식 (2)}$$

식 (2)에서 S는 척도인자(scale factor)로, 두 영상 좌표계의 척도를 맞추는데 필요하다. 최종적으로 원본영상에서 식 (2)의 계산을 통해 원근현상이 완화된 변환영상을 얻을 수 있다.

## 2.2 딥러닝 기반 객체인식 모델

본 논문에서는 변환영상의 원근현상의 완화를 통해 터널 영상유고 시스템의 성능을 향상할 수 있음을 보이기 위해 딥러닝 객체인식 모델의 학습 및 추론을 통한 비교 검토 연구를 수행하였다.

딥러닝은 기계학습의 종류 중 하나로, 인간의 명령 없이 컴퓨터가 데이터의 학습을 통해 자동으로 모델을 개선할 수 있으며, 기존 기계학습보다 더 깊고 복잡한 구조를 가진다.

객체인식은 목표 대상 객체를 직사각형 형태의 경계박스를 검출하는 기술이며, 터널 영상유고 시스템에서는 정차 및 역주행 유고상황을 탐지하기 위해 차량 객체가 필요하다. 그리고 차량을 포함하여 보행자 및 화재 객체 3종류의 객체를 검출하는 것이 필요하다.

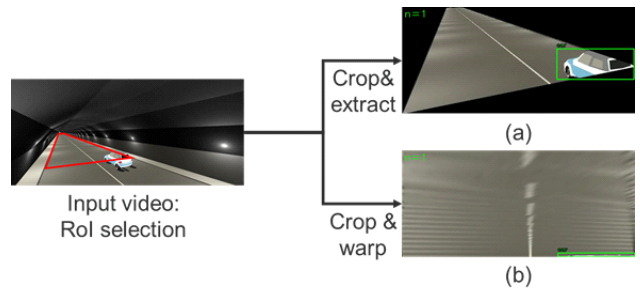
한편, 딥러닝 객체인식 모델은 영상과 경계박스 및 객체의 종류 데이터를 학습함으로써 스스로 주어진 영상에서 목표 객체를 찾는 것이 가능하다. 그런데, 데이터의 차원을 줄이

면서 특징만을 남기는 딥러닝 모델 특성으로 인해 객체의 크기가 작으면 객체의 특징을 남기기 어려우므로 학습 및 추론이 어렵다. 즉, 원본영상에서 터널 CCTV와 멀리 떨어진 객체는 영상에서 보이는 크기가 매우 작으므로 딥러닝 모델의 학습이 쉽지 않을 것이다. 반면에 변환영상은 먼 거리에 존재하는 객체의 크기가 확대되었으므로 딥러닝 모델의 학습 및 추론이 원본영상보다 수월하게 된다.

## 3. 실험

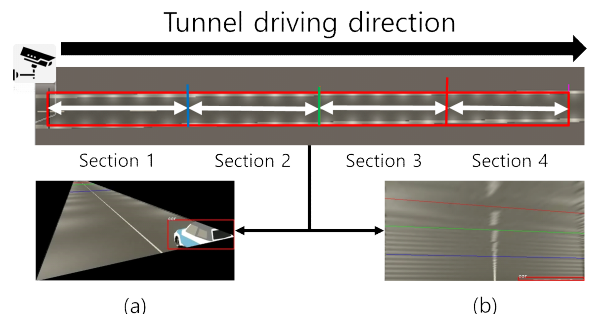
### 3.1 실험 조건

본 논문에서는 변환영상의 원근현상의 완화를 통해 터널 영상유고 시스템의 성능이 향상될 수 있음을 보이기 위해 딥러닝 객체인식 모델의 학습을 통한 비교 검토 연구를 수행하였다. 그림 1과 같이 터널 CCTV 영상에서 실제 200m 주행방향 거리만큼 관심영역을 설정하였다. 다음, 그림 1(a)와 같이 관심영역 밖의 영역은 검정색 배경으로 처리하여 case 1으로 명명하였다. 이는 원본영상 관심영역을 의미한다.



[그림 1] 역 원근변환 효과를 검토하기 위한 case별 영상 (a) 원본영상(case 1) (b) 변환영상 (case 2)

그리고 그림 1(b)와 같이 역 원근변환을 통해 얻은 변환영상은 case 2로 명명하였으며, 두 개의 case들에 대해 차량 객체만을 대상으로 하여 데이터셋 레이블링을 진행하였다. 각 데이터셋은 동일한 영상을 기반으로 제작되었으므로 동일한 정지영상 개수와 동일한 객체수를 가지며, 학습용 데이터셋과 추론용 데이터셋을 나눌 때 동일한 정지영상 조건으로 나누어진다.



[그림 2] 각 영상 조건의 관심영역에서 각각 50m 간격으로 구분된 영상모델(a) case 1 (b) case 2

그리고 데이터셋을 추론할 때, 실제로 거리에 따른 객체 크기 차이에 따라 딥러닝 모델의 객체인식 성능에 차이가 발생하는지 확인할 필요가 있으므로 그림 2와 같이 주행방향 200m의 관심영역 거리범위에서 50m 간격으로 4개의 구간을 나누는 다음, 해당 범위 안에 있는 객체만을 포함한 데이터셋을 제작하였다. 각 구간에서 객체를 포함하는 기준은 x축을 수평축, y축을 수직축이라고 할 때, 경계박스의 y 최소값과 y 최대값의 평균을 기준으로 구분하였다.

상기 언급한대로 case마다 데이터셋을 제작하였으며, 그 현황을 표 1과 같다. case 1은 원본영상을 사용한 경우, case 2는 변환영상을 사용한 경우이다.

[표 1] 딥러닝 객체인식 데이터셋 현황

Case	데이터셋 종류	실험 구간	정지영상 개수	차량 객체수
case 1	train	-	421	1387
	test	section 1	58	111
		section 2	55	87
		section 3	63	83
		section 4	70	92
		총계	106	373
case 2	train	-	421	1387
	test	section 1	59	104
		section 2	63	91
		section 3	62	83
		section 4	67	95
		총계	106	373

두 case 모두 영상의 폭과 높이는 각각 646, 324로 동일하며, 데이터셋은 딥러닝 객체인식 모델의 학습만을 위한 학습 데이터셋과 검증만을 위한 검증(test) 데이터셋으로 나누었으며, 정지영상 개수 기준 각각 8:2의 비율로 나누었다. 본 논문에서는 두 case의 비교는 검증 데이터셋에서만 비교를 진행할 것이므로 구간 분할은 test 데이터셋에서만 진행하였다.

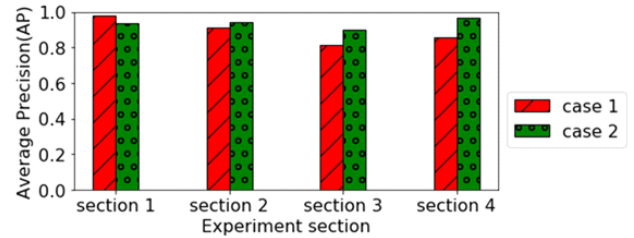
딥러닝 객체인식 모델은 Faster R-CNN을 사용하며, learning rate는 0.0001, 정지영상에 대한 batch size는 1, epoch는 100으로 설정하여 학습을 진행하였다. Convolutional layer는 50층으로 구성된 ResNet을 사용하였다.

그리고 AMD RYZEN 1800X, NVIDIA GTX 1080TI 11GB 사양을 가진 컴퓨터와 python 3.7, tensorflow 1.15.5 버전 소프트웨어로 딥러닝 모델의 학습을 진행하였다.

### 3.2 실험 결과 및 분석

딥러닝 객체인식 모델의 성능 평가는 객체인식 분야에서 널리 활용되고 있는 지표인 Average Precision (AP)을 활용하였다. 같은 객체에 대하여 레이블링된 경계박스와 추론된 경계박스를 비교할 때, 두 경계박스간 겹치는 비율을 말하는 IoU(intersection over union)를 기준으로 참 또는 거짓을 판별하였으며, 본 논문에서는 IoU 문턱값을 0.5로 설정하였다.

그림 3과 같이 본 논문에서는 두 case에 대해 학습에 사용되지 않은 검증 데이터셋에 대한 추론 결과를 4개의 구간마다 AP값을 비교한 그래프이다. 그림 5의 section 1에서 case 1의 AP값이 case 2보다 더 높는데, 터널 CCTV와 가까운 지점에서는 원본영상에서 보이는 객체의 크기가 매우 크기 때문에 딥러닝 객체인식모델이 학습하기에 유리함을 알 수 있다.



[그림 3] 두 case 및 실험 구간에 따른 test 데이터셋에 대한 AP값 비교결과 그래프

그런데, section 2에서는 두 case가 비슷한 AP값을 가지며, section 3와 4에서는 case 1의 AP값이 case 2보다 확연하게 떨어진다. 이러한 경향은 변환영상에서 터널 CCTV와 먼 지점의 차량 객체의 크기가 확대됨으로 인해 딥러닝 모델의 학습이 수월했음을 알 수 있다.

### 4. 결론

본 논문에서는 터널 CCTV에서 발생하는 심한 원근현상을 극복하기 위해 역 원근변환 기법을 도입하였으며, 기존 원본영상과 성능차이가 발생함을 확인하기 위해 동일한 조건의 관심영역을 설정한 다음, 원본영상과 변환영상에 대한 데이터셋을 제작하였고, 각각에 대한 딥러닝 모델에 대하여 동일한 조건으로 딥러닝 모델을 학습한 결과, 다음과 같은 결론을 얻을 수 있었다.

터널 영상유고 시스템을 구성할 때, 터널 CCTV 간격이 50m이하인 경우에는 변환영상을 사용하지 않는 것이 더 나은 선택일 수 있다. 그러나 터널 CCTV 간격이 50m를 초과하는 현장에서는 변환영상을 사용하는 것이 유리하며, 대부분의 터널 현장이 도로터널 방재시설 설치 및 관리지침에 의거하여 200m~400m의 간격으로 설치되어 있으므로 변환영상을 활용한 터널 영상유고 시스템이 효과적으로 유고상황을 검출

할 수 있을 것이다.

또한, 변환영상에서의 객체인식에서는 가깝고 먼 동일객체의 객체인식 경계박스의 크기가 유사하므로, 추론된 경계박스 이동 정보를 기반으로 계산되는 유고상황(역주행, 정차 등) 평가 알고리즘에는 매우 유리하다 할 수 있다.

#### 참고문헌

- [1] 국토교통부, “도로터널 방재시설 설치 및 관리지침,” 2020년.
- [2] G. Salton and M. J. McGill, “Introduction to modern information retrieval,” McGraw-hill, Inc., NY, USA, 1983.
- [3] Z. Liu, S. Wang and X. Ding, “ROI perspective transform based road marking detection and recognition,” Proceedings of 2012 International Conference on Audio, Language and Image Processing, pp. 841-846, 2012.
- [4] R. Pflugfelder, H. Bischof, G. F. Dominguez, M. Nolle and H. Schwabach, “Influence of Camera Properties on Image Analysis in Visual Tunnel Surveillance,” Proceedings of ‘2005 IEEE Intelligent Transportation Systems’, pp. 868-873, 2005.
- [5] L. Quan and R. Mohr, “Determining perspective structures using hierarchical Hough transform,” Pattern Recognition Letters, vol. 3, no. 4, pp. 279-286, 1989.
- [6] S. Ren, K. He, R. Girshick and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 6, pp. 1137-1149, 2016.
- [7] Y. Xu, S. Mita, H. T. Niknejad and K. Ishimaru, “Real-time Stereo Vision System at Tunnel,” Proceedings of International Conference on Computer Vision Theory and Applications, pp. 402-409, 2017.