

역 원근변환 기법을 이용한 터널 영상유고시스템의 원거리 감지 성능 향상에 관한 연구

이규범¹ · 신휴성^{2*}

¹학생회원, 한국건설기술연구원 미래스마트건설연구본부 학생연구원, 과학기술연합대학원대학교(UST) 스마트도시건설융합 석박통합과정
²정회원, 한국건설기술연구원 미래스마트건설연구본부 선임연구위원

A study for improvement of far-distance performance of a tunnel accident detection system by using an inverse perspective transformation

Kyu Beom Lee¹ · Hyu-Soung Shin^{2*}

¹Student Researcher, Dept. of Future & Smart Construction Research, Korea Institute of Civil Engineering and Building Technology
Ph.D. Student, Smart City and Construction Convergence, University of Science & Technology
²Senior Research Fellow, Dept. of Future & Smart Construction Research, Korea Institute of Civil Engineering and Building Technology

*Corresponding Author : Hyu-Soung Shin, hyushin@kict.re.kr

OPEN ACCESS

Journal of Korean Tunnelling and Underground Space Association
24(3)247-262(2022)
<https://doi.org/10.9711/KTAJ.2022.24.3.247>

eISSN: 2287-4747
pISSN: 2233-8292

Received March 10, 2022
Revised May 12, 2022
Accepted May 18, 2022



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2022, Korean Tunnelling and Underground Space Association

Abstract

In domestic tunnels, it is mandatory to install CCTVs in tunnels longer than 200 m which are also recommended by installation of a CCTV-based automatic accident detection system. In general, the CCTVs in the tunnel are installed at a low height as well as near by the moving vehicles due to the spatial limitation of tunnel structure, so a severe perspective effect takes place in the distance of installed CCTV and moving vehicles. Because of this effect, conventional CCTV-based accident detection systems in tunnel are known in general to be very hard to achieve the performance in detection of unexpected accidents such as stop or reversely moving vehicles, person on the road and fires, especially far from 100 m. Therefore, in this study, the region of interest is set up and a new concept of inverse perspective transformation technique is introduced. Since moving vehicles in the transformed image is enlarged proportionally to the distance from CCTV, it is possible to achieve consistency in object detection and identification of actual speed of moving vehicles in distance. To show this aspect, two datasets in the same conditions are composed with the original and the transformed images of CCTV in tunnel, respectively. A comparison of variation of appearance

speed and size of moving vehicles in distance are made. Then, the performances of the object detection in distance are compared with respect to the both trained deep-learning models. As a result, the model case with the transformed images are able to achieve consistent performance in object and accident detections in distance even by 200 m.

Keywords: Tunnel CCTV-based accident detection system, Inverse perspective transform, Perspective, Transformed image, Deep learning

초 록

국내 200 m 이상 연장의 터널에서는 CCTV 설치가 의무화되어 있으며, 터널 내 돌발 상황을 자동으로 인지한 다음 터널 관리자에게 알릴 수 있는 터널 영상유고시스템의 운영이 권고된다. 여기서 터널 내 설치된 CCTV는 터널 구조물의 공간적인 한계로 인해 낮은 높이로 설치된다. 이에 따라 이동차량과 매우 인접하므로, 이동차량과 CCTV와의 거리에 따른 원근현상이 매우 심하다. 이로 인해, 기존 터널 영상유고시스템은 터널 CCTV로부터 멀리 떨어질수록 차량의 정차 및 역주행, 보행자 출현 및 화재 발생과 같은 터널 내 유고상황을 인지하기 매우 어려우며, 100 m 이상의 거리에서는 높은 유고 상황 인지 성능을 기대하기 어려운 것으로 알려져 있다. 이 문제를 해결하기 위해 관심영역 설정 및 역 원근변환(Inverse perspective transform)을 도입하였으며, 이 과정을 통해 얻은 변환영상은 먼 거리에 있는 객체의 크기가 확대된다. 이에 따라 거리에 따라 객체의 크기가 비교적 일정하게 유지되므로, 거리에 따른 객체 인식 성능과 영상에서 보이는 차량의 이동속도 또한 일관성을 유지할 수 있다. 이를 증명하기 위해 본 논문에서는 터널 CCTV의 원본영상과 변환영상을 바탕으로 동일한 조건을 가지는 데이터셋을 각각 제작 및 구성하였으며, 영상 내 차량의 실제 위치의 변화에 따른 겉보기 속도와 객체 크기를 비교하였다. 그 다음 딥러닝 객체인식 모델의 학습 및 추론을 통해 각 영상 데이터셋에 대한 거리에 따른 객체인식 성능을 비교하였다. 결과적으로 변환영상을 사용한 모델은 200 m 이상의 거리에서도 객체인식 성능과 이동 차량의 유고상황 인지 성능을 확보할 수 있음을 보였다.

주요어: 터널 영상유고시스템, 역 원근변환, 원근현상, 변환영상, 딥러닝

1. 서론

도로 터널에서 발생하는 돌발사고를 의미하는 유고상황은 대피공간의 부족, 낮은 조도 등과 같은 열악한 터널 환경 조건으로 인해 외부 도로 환경보다 사고 피해 정도가 크고, 중대형 2차 사고로 이어질 가능성도 매우 크다. 따라서 운용 중인 터널을 관리하는 주체는 유고상황을 조기에 정확히 인지한 다음 신속하게 대응하는 것이 중요하다. 이에 대한 대응책으로 국토교통부에서는 방재등급 3등급 이상 터널에 대하여 CCTV의 설치를 의무화 및 감시하도록 하였으며, 방재등급이 3등급 미만인 경우, 200 m 이상의 연장을 가진 터널에서는 CCTV의 설치를 권고한다. 그리고 터널을 관리하는 주체를 보조하기 위해 유고상황을 자동으로 인지 및 알릴 수 있는 터널 영상유고시스템의 설치 및 운영을 지침으로 권고하고 있다(MOLIT, 2021).

한편 터널 CCTV는 외부 일반 도로에 설치된 CCTV와 비교했을 때, 낮은 조도와 짙은 먼지로 인해 선명한 영상을 확보하는 것이 어려우며, 공간적인 한계로 인해 터널 내 CCTV 설치 높이를 충분히 확보하기 어려우므로 이동

차량 측선과 매우 가깝게 설치될 수밖에 없다. 이에 따라 CCTV 영상에 보이는 이동차량은 거리에 따라 극심한 원근현상이 발생하게 된다.

두 환경을 비교했을 때, 외부 일반 도로에 설치된 CCTV는 공간상 제약이 없으므로 CCTV를 높게 설치할 수 있으므로 원근현상을 줄이는데 유리하며, 국내에서 일반 도로에 설치된 CCTV는 최소 15 m 이상 높이에서 CCTV를 설치하도록 되어있다(MOLIT, 2010).

하지만 터널은 공간상 제약으로 인해 터널 CCTV의 설치 높이가 한정된다. 이에 따라 터널 CCTV 영상은 정면에서 보는 정면도와 유사해지므로 원근현상이 심해진다. 실제로 경기도의 76개 도로터널에 대한 평균 높이는 6.04 m이고, 규정상 터널 내 CCTV 최소 설치 높이는 3.5 m 이상이므로 일반 도로에 대한 CCTV 설치 높이보다 낮은 것을 확인할 수 있다(Gyeonggi Province, 2022).

이러한 이유로 외부 환경에 대응하여 개발된 영상유고시스템은 터널 내 설치 및 운용 시, 터널 내 CCTV와 먼 지점에 대한 객체 인식 및 유고상황 인지 성능이 현격히 떨어진다. 따라서 국토교통부에서는 터널 영상유고시스템이 설치된 경우, 터널 CCTV와 먼 지점에서 영상유고시스템의 인지 성능을 담보할 수 없기 때문에 터널 내 CCTV의 설치 간격을 규정된 CCTV 설치간격인 200~400 m보다 조밀한 100 m 이내의 간격으로 설치하도록 규정하고 있다(MOLIT, 2021).

터널 영상유고시스템의 인지성능 발현의 핵심 요소는 객체인식 성능이다(Shin et al., 2017b). 하지만 상기 설명한 심한 원근현상으로 인해 터널 CCTV로부터 멀어지는 차량 객체는 크기가 급격히 작아지며 객체인식 성능도 현격히 저하된다(Tong et al., 2020). 이는 결과적으로 터널 영상유고시스템의 인지성능 하락으로 이어진다(Pflugfelder et al., 2005).

이 문제를 극복하기 위해 Min 등(Min et al., 2019)은 터널 내 CCTV 영상을 고해상도로 재현시켜 보행자의 객체인식 성능을 향상시키는 연구를 진행하였으며, 기존 객체인식 알고리즘 대비 보행자에 대한 객체인식 성능을 88%에서 90%로 2% 정도 향상시킨 것으로 보고한 바 있다. 그러나 고해상도 영상을 바탕으로 객체인식을 진행할 경우, 객체인식 속도가 현저히 느려진다. 터널 영상유고시스템은 신속한 유고상황 인지를 위해 정확함뿐만 아니라 빠른 객체인식을 요구하므로, 고해상도로 영상을 재현시키는 방법은 시스템의 현장 적용성 및 효율성 측면에서 적절한 방법이 아니다.

한편 터널 CCTV영상은 낮은 조도로 인해 터널 내 조명에 민감하게 반응하므로, 객체인식 성능에 영향을 미치게 된다. 이 문제를 해결하기 위해 Tsai 등(Tsai et al., 2016)은 영상에서 조도 보상(Illumination compensation)을 통해 터널 내 조명의 영향을 줄이는 방법을 제안한 바 있다. 이를 통해 기존 터널 CCTV 영상에서는 객체인식 알고리즘이 차량 객체를 하나의 형태로 인식하지 못하였지만, 조도 보상을 진행한 영상에서는 차량 객체를 하나의 형태로 인식하는 것이 가능하였다. 그럼에도 불구하고 터널 CCTV와 멀리 떨어질 때, 객체인식 알고리즘은 터널 배경을 차량으로 인식하는 경우가 발생하므로 심한 원근현상에 대한 근본적인 해결책이 되지 못한다.

따라서 본 연구에서는 멀리 있는 객체의 크기를 가까이 있었을 때와 유사하게 만들어 줌으로써, 거리에 따른 객체의 크기에 대한 민감도를 최소화시켜줄 수 있는 역 원근변환(Inverse perspective transform) 기법을 도입하였다.

이어 가상의 터널 모델링 및 이동차량에 대한 영상 데이터셋을 구축한 다음 다양한 실험을 통해 도입한 기법의 타당성과 터널 영상유고시스템의 인지성능 향상 효과를 정량적으로 파악하기 위한 연구를 진행하였다.

2. 역 원근변환 기법

역 원근변환 기법은 영상을 구성하는 픽셀의 배치 구조를 변경함으로써, 일정한 법칙에 따라 영상을 기하학적으로 변환시키는 방법을 말한다. 역 원근변환 기법은 영상변환 방법 중 하나로, 주어진 영상에서 원근현상을 보정하여 완화시킬 수 있는 영상변환 방법이다(Mallot et al., 1991). 본 기법은 현재 자율주행 분야에서 차선인식(Bertozz et al., 1998) 및 주변 방해물 인식과 차량 인식(Lee et al., 2019) 등에 사용되고 있으며, CCTV를 기반으로 실시간 끼어들기 위반차량을 검출하는 시스템에도 적용되고 있다(Lee et al., 2011). 본 논문에서는 터널 CCTV 영상에서 나타나는 심한 원근현상의 완화를 통한 터널 영상유고시스템의 유고상황 인지성능 향상을 위해 본 기법을 도입하였다.

2.1 가상 터널 모델링 및 좌표계 묘사

본 논문에서는 역 원근변환의 도입에 대한 필요성을 보이기 위해 원근현상의 가시적 특징에 대해 고찰하였다. 실제 터널 CCTV를 바탕으로 실험적 고찰을 진행하기 위해서는 실제 도로터널 현장에 대한 거리의 실측 및 차량의 위치정보 측정이 필요하다. 하지만 도로 터널은 타 차량의 주행으로 인해 실측 작업에 위험성이 존재하며, 현장에서 주행 중인 차량의 주행속도가 일정하지 않는 문제가 있다. 따라서 본 논문에서는 유니티 엔진 게임 개발 플랫폼 기반으로 가상 터널을 모델링 하였으며(Juliani et al., 2018), 가상터널에서 CCTV의 설치 및 차량의 촬영을

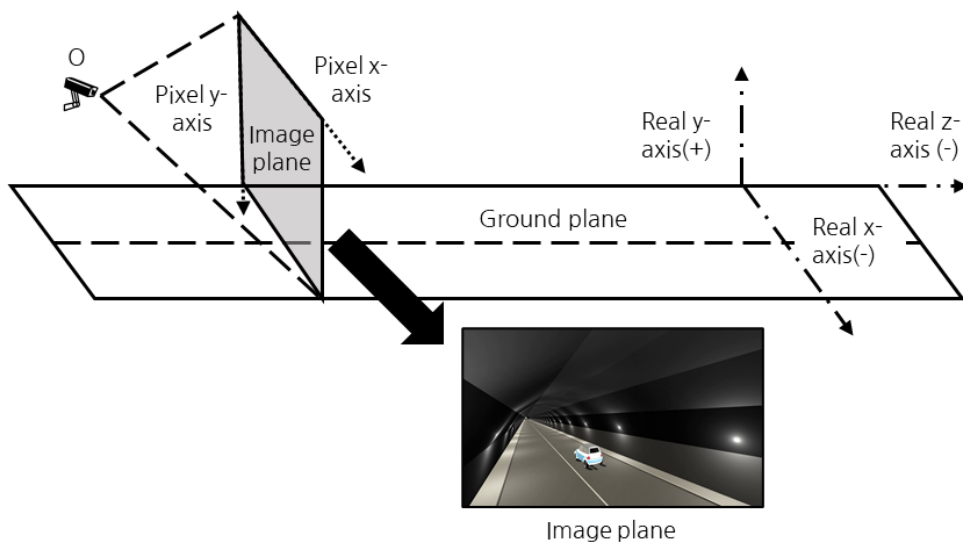


Fig. 1. Real coordinate system and image coordinate system in virtual tunnel model

통해 터널 CCTV에서 발생하는 원근현상의 특징과 문제점을 파악하였다.

가상 터널에서 실제 좌표계와 터널 CCTV에서 촬영된 영상 좌표계의 조건은 Fig. 1과 같이 표현할 수 있다. 실제 좌표계에서 터널 CCTV 영상은 ‘O’ 지점에서 촬영되며, 차량은 도로 지표면(Ground plane) 선상을 정속으로 주행한다. 실제 현장의 좌표계는 지표면을 기준으로 차량이 주행하는 방향을 음의 방향으로 z축, 높이 방향은 y축, 차량이 주행하는 방향의 좌우방향이 x축이다. 한편 터널 CCTV에서 촬영된 영상은 2차원의 영상 좌표계로 표현할 수 있으며, 영상면(Image plane)에서 수평방향은 픽셀 x축, 수직방향은 픽셀 y축으로 정의할 수 있다.

2.2 터널 CCTV에서 관심영역 설정조건

터널 CCTV 영상(이하 원본영상)에서 역 원근변환 기법을 적용하려면 관심영역(Region of interest, ROI)을 설정해야 한다. 관심영역은 Fig. 2(a)와 같이 수평면 영상(Top-view image)으로 펼쳐 확인하였을 때, 도로 폭 및 CCTV 설치간격에 맞추어 직사각형 모양으로 표현되어야 한다. 이를 위해서는, Fig. 2(b)의 원본영상에서 다음의 절차에 따라 관심영역을 설정해야 한다.

- (1) 원본영상에서 차량 주행방향과 평행한 좌우측 최외곽 차선 상에 직선 두 개를 그린다.
- (2) 실제 좌표계에서 터널 CCTV와 가장 가까우며, 차량 주행방향과 수직인 직선과 (1)에서 구한 두 직선 사이 교점 2개를 구한다.
- (3) 실제 좌표계에서 목표한 최대 인지 지점에서 차량 주행방향과 수직인 직선과 (1)에서 구한 두 직선 사이 교점 2개를 구한다.

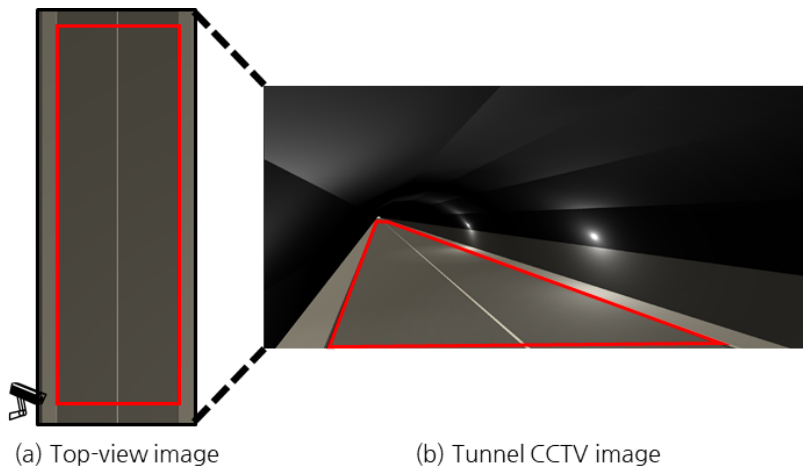


Fig. 2. Condition for setting a region of interest in the tunnel CCTV image

이러한 절차를 통해 관심영역을 정의하는 4개의 점을 설정할 수 있다. 설정된 관심영역 4개의 꼭지점 좌표를 이용하여, 원본영상에서 수평면도 상으로 영상변환을 수행한다. 수평면도 영상으로 변환된 관심영역은 직사각형이다. 그리고 영상변환을 수행하기 위해서 원본영상과 수평면도 상으로 변환된 영상의 동일한 두 점들 간의 좌표 변환 관계를 표현하는 호모그래피(Homography) 행렬이 필요하며(Galeano et al., 2011), 다음 식 (1)과 같이 표현할 수 있다.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \tag{1}$$

식 (1)에서 H는 호모그래피 행렬을 말하며, 호모그래피 행렬의 각 값들은 좌표변환 관계를 표현하는 값들이다. 역 원근변환은 원본영상 좌표계의 점 (x,y)가 있고, 변환영상 좌표계의 점 (x',y')이 있을 때, 다음 식 (2)와 같이 역 원근변환을 통해 임의의 사각형인 관심영역에서 직사각형으로 기하학적 좌표계 변환이 가능하다. 두 좌표계의 단위는 픽셀이다.

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

식 (2)에서 S는 척도인자(Scale factor)로 두 좌표계의 척도를 맞추는데 필요한 값이다. 결과적으로 역 원근변환은 원본영상에서 식 (2)를 적용해 변환하면 Fig. 2(a)와 같이 수평면도처럼 보이는 변환영상(Transformed image)을 얻을 수 있다.

그 다음 터널 영상유고시스템의 유고상황 인지영역을 의미하는 관심영역은 원본영상과 변환영상 모두 동일한 영역으로 설정해야 한다. 따라서 기 설정된 직사각형의 변환영상 관심영역을 최종 결정하기 위한 후처리 프로세스를 수행해야 한다.

Fig. 3은 기 설정된 관심영역 및 호모그래피 행렬을 이용한 역 원근변환을 실행하는 과정을 보여준다. 후처리 프로세스는 먼저 Fig. 3(a)와 같이 원본영상에서 관심영역을 설정한 다음 호모그래피 행렬의 구성 및 역 원근변환을 실행하여 Fig. 3(b)와 같이 원본영상과 대응되는 변환영상을 얻는다. Fig. 3(b)의 변환영상의 해상도는 원본영상과 동일하다. 그 다음 Fig. 3(c)와 같이 원본영상의 관심영역과 대응되는 직사각형의 관심영역을 변환영상에서 설정한다. 마지막으로 변환영상에서 직사각형의 관심영역을 추출하여, Fig. 3(d)와 같이 원근현상이 완화된 변환영상을 얻을 수 있다.

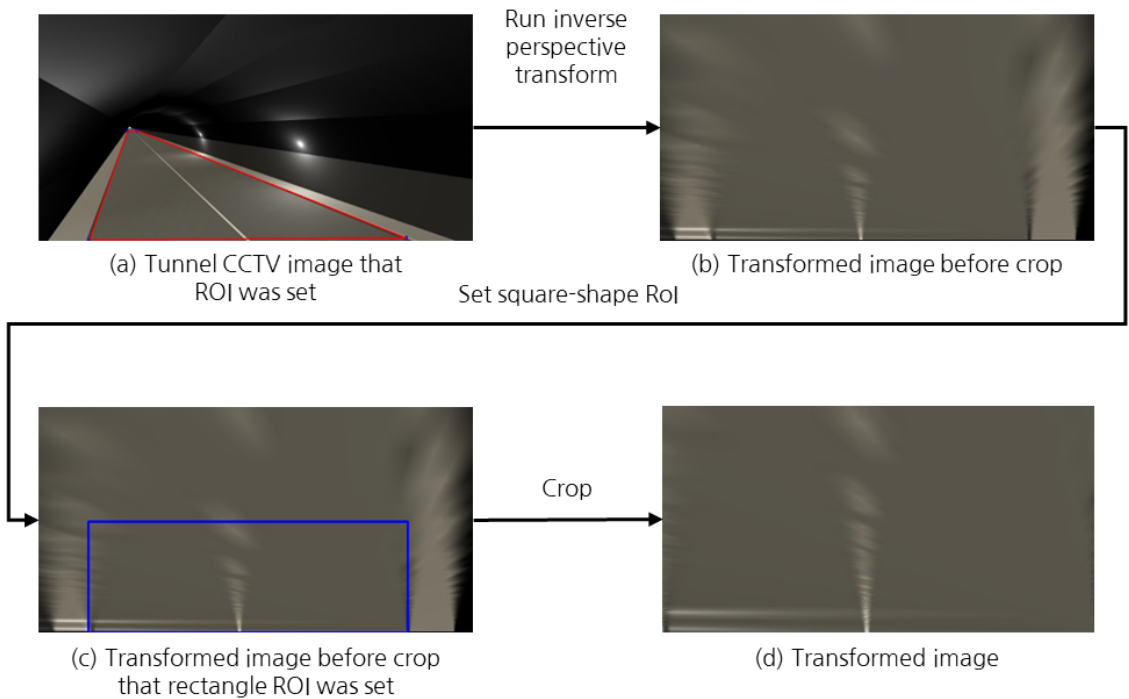


Fig. 3. The process of obtaining a transformed image

3. 원근현상 저감효과 비교 검토 실험

3.1 실험 개요

본 논문에서는 원근현상 완화에 대한 터널 영상유고시스템의 성능 검토를 위해 원본영상과 변환영상 내 차량의 위치이동에 따른 겉보기 속도와 차량객체 크기의 변화를 고찰하였다. 그 다음 각 영상에서 딥러닝 학습모델의 객체인식 성능을 비교 검토하였다. 본 실험은 원본영상과 변환영상의 실험조건이 같아야 한다. 따라서 두 영상의 관심영역의 절대 면적과 크기가 동일하며, Fig. 4와 같이 차량 주행방향으로 터널 내 관심영역 설정범위를 일반 터널 CCTV 설치 간격인 200 m로 설정하였다. 그리고 원본영상에 해당하는 실험조건은 Fig. 4(a)와 같이 관심영역 밖의 배경에 대해 검정색 배경으로 처리한 다음 자르기(Crop)를 진행하여 얻은 영상에 대해 Case 1으로 명명하였다(Shin et al., 2017a). 그리고 변환영상에 해당하는 실험조건은 위 2.3절에서 설명한 방법으로 기 설정된 관심영역에서 호모그래피 행렬을 찾은 다음 Fig. 4(b)와 같이 역 원근변환 및 자르기를 진행하여 얻은 영상을 말하며 Case 2로 명명하였다.

그 다음 각 Case에 대한 데이터셋은 동일한 영상을 기반으로 레이블링이 진행되었다. 따라서 2개의 데이터셋은 동일한 정지영상 개수와 동일한 객체 수를 가지며, 학습용 데이터셋과 추론용 데이터셋 구분도 동일하게 진행하였다.

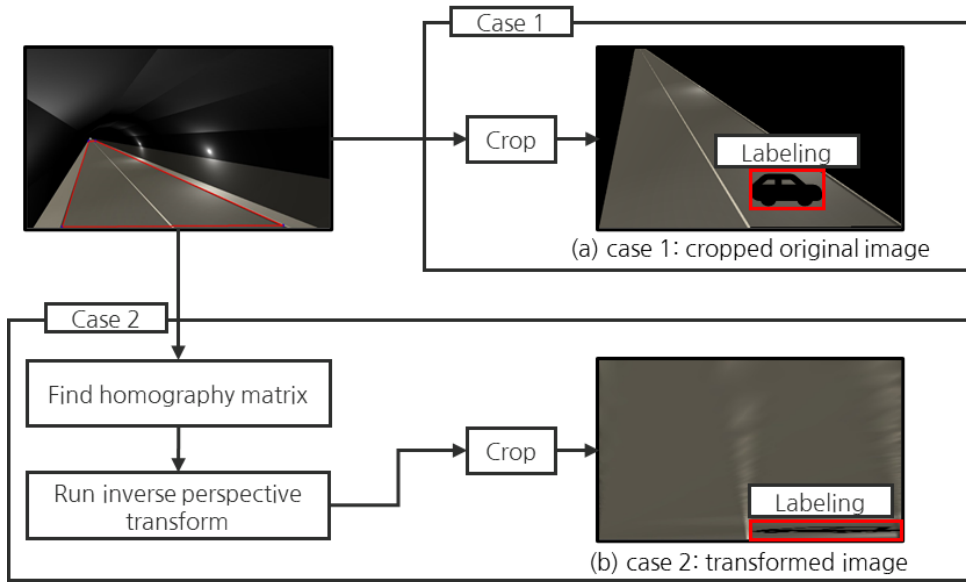


Fig. 4. Experimental conditions divided according to the ROI set in the tunnel CCTV image

3.2 영상 내 보여지는 객체에 대한 원근현상의 영향 검토

3.2.1 실제 위치변화에 따른 겉보기 속도 변화

겉보기 속도(Apppearance velocity)는 영상에서 보여지는 속도로 정의한다. 원근현상이 심한 경우, 실제속도는 일정하지만 겉보기 속도는 급격하게 감소하게 된다. 여기서 본 논문에서 제안한 역 원근변환을 통해 원근현상을 완화하면, 겉보기 속도도 실제속도와 유사하게 변환시킬 수 있다.

Fig. 5는 겉보기 속도를 계산하는 방법을 나타낸 것이다. 겉보기 속도는 영상 좌표계에서 계산되며, 터널의 주행방향은 실제 현장 좌표계에서는 z방향, 영상에서는 Fig. 5와 같이 y방향이다. 차량의 탐지 시작점인 0을 기준으로 주행방향으로 음의 거리로 나타낸다. 변환영상에서는 수직방향 픽셀값의 변화량을 dy , 시간차를 dt 로 정의하고 겉보기 속도 V 를 다음 식(3)과 같이 나타낼 수 있다. 식(3)에서 겉보기 속도를 의미하는 V 의 단위는 픽셀/초로 나타낸다.

$$V = \frac{dy}{dt} \tag{3}$$

본 실험에서는 차량을 주행방향으로 시속 60 km 정속 주행시켰다. 차량의 실제 위치에 따른 겉보기 속도의 변화는 Fig. 6과 같다.

Fig. 6에서 원본영상인 Case 1에서는 원근현상이 심하기 때문에 터널 CCTV와 멀어질수록 겉보기 속도가 급격하게 감소하는 것을 확인할 수 있으며, 시작선 상에서 100 m 떨어진 지점을 지나갈 때, 차량의 겉보기 속도는 0에

가까워지는 경향을 보인다. 한편 변환영상인 Case 2에서는 실제 위치와 상관없이 거의 일정한 겉보기 속도를 유지하는 경향을 보인다.

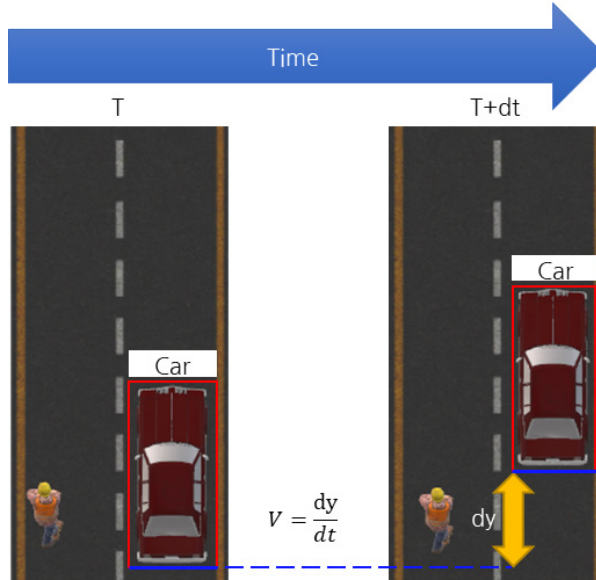


Fig. 5. Appearance velocity calculation process for a single vehicle object

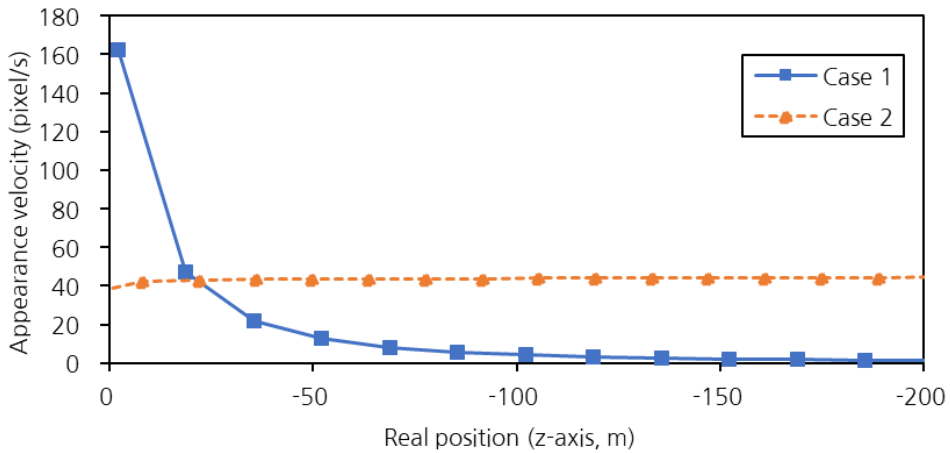


Fig. 6. Comparison of the appearance velocity of the single vehicle as the real position was changed

3.2.2 실제 위치변화에 따른 차량 객체의 크기 변화

원근현상 완화는 영상에서 보여지는 차량 객체의 실제 위치에 따른 크기 변화에서도 크게 영향을 미친다. 차량 객체의 크기는 차량 경계박스에서 수평축인 x축 방향으로 경계박스의 폭, y축 방향으로 경계박스의 높이를 정의

할 수 있다.

Fig. 7은 원본영상과 변환영상 각각에서 실제 위치에 따른 차량 객체의 폭과 높이의 변화를 보여준다. 터널 CCTV는 0 m의 위치에 존재하며, 차량은 실제 좌표계에서 z축 음의 방향으로 진행하게 된다. Fig. 7(a)에서 차량 객체의 폭을 먼저 비교해보면 원본영상인 Case 1은 터널 CCTV와 차량간 거리가 멀어질수록 객체의 폭이 급격하게 감소하지만, Case 2는 비교적 일정한 크기로 유지됨을 확인할 수 있다. 한편 Fig. 7(b)에서 객체의 높이 변화를 비교했을 때, Case 1은 Fig. 7(a)와 마찬가지로 객체의 높이가 급격하게 감소한다. 그런데 Case 2는 차량의 진행에 따라 객체의 높이가 증가하다가, 객체의 위치가 -150 m 지점일 때부터 감소하는 경향을 보인다.

일반적으로 차량 객체가 출현할 때와 나갈 때는 관심영역 기준으로 차량의 상단부터 부분적으로 보이며 들어오고, 상단부터 사라지며 관심영역을 벗어나게 된다. 원본영상에서는 원근현상으로 인해 이러한 현상을 관찰하기 힘들지만 변환영상에서는 객체의 크기가 비교적 일정하게 유지되므로 이 현상을 확인할 수 있다.

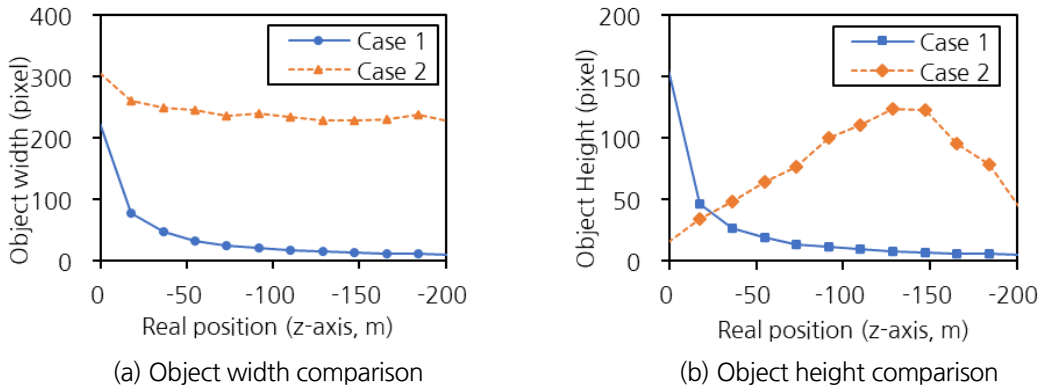


Fig. 7. Comparison of the object width and height of the single vehicle as the real position was changed

3.3 딥러닝 객체인식 모델 학습을 통한 원근현상 영향 검토

3.3.1 개요

본 논문에서 터널 영상유고 감지를 위해 사용한 딥러닝 객체인식 모델은 해당 분야에서 표준이 되는 Faster R-CNN (Faster region-based convolutional neural network)을 사용한다(Ren et al., 2015). Faster R-CNN 모델의 구조는 Fig. 8과 같으며 정지영상(Image)을 입력값으로 한다. 그 다음 합성곱 신경망층(Convolutional layer)에서 특징도를 추출하며, 특징도를 기반으로 영역 제안 신경망층(Region proposal network)에서 정지영상 내 객체가 있을 법한 위치에서 경계박스를 제안한다. 마지막으로 신경망층(Neural network layer)은 특징도와 제안된 경계박스들을 바탕으로 객체의 종류(Object class)와 정확한 경계박스 위치를 결정한다.

한편 현존하는 딥러닝 객체인식 모델의 핵심요소는 합성곱 신경망층이며, 정지영상에서 특징적인 요소만 추출된 특징도(Feature map)를 얻을 수 있다(LeCun et al., 1990; 2015; Zou et al., 2019). 여기서 특징도의 해상도는

입력값으로 사용된 원본영상보다 줄어들며, 이 특징으로 인해 딥러닝 객체인식 모델은 객체의 크기가 작으면 객체의 특징을 남기기 어려우므로 학습 및 추론이 상대적으로 어려워진다(Eggert et al., 2017). 이러한 특성은 첫번째 실험에서 터널 CCTV와 멀어질수록 Case 1의 차량 객체 크기가 작아지는 것을 감안했을 때, Case 1에 대한 딥러닝 모델의 객체인식 성능이 상대적으로 떨어질 것이라 추정할 수 있다. 반면에 Case 2는 먼 거리에 존재하는 객체의 크기가 확대되므로, 딥러닝 모델의 객체인식 성능이 Case 1보다 높을 것으로 기대할 수 있다.

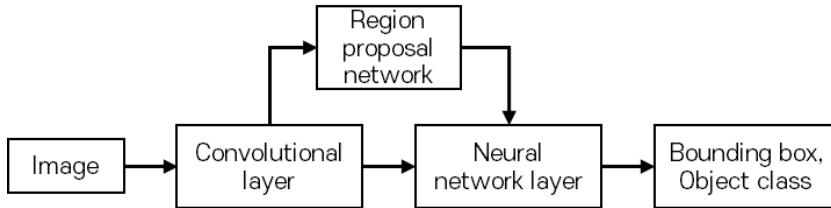


Fig. 8. Simple structure of faster R-CNN

3.3.2 실험 방법

본 논문에서는 동일한 조건에서 동일한 객체에 대하여 Case 1과 Case 2를 각각 학습한 딥러닝 모델에 대하여 객체인식 성능 지표인 평균 정밀도(Average precision, AP)라는 지표를 통해 비교하였다(Zhu, 2004). 또한 Case 1과 Case 2의 객체인식 추론성능 비교 검토를 통해 동일한 CCTV에서 터널 영상유고시스템의 영상유고 가탐거리에 대해 고찰하였다. 따라서 두 모델에 대한 AP값 비교는 터널 CCTV와 차량 객체간 거리에 따라 진행할 것이다.

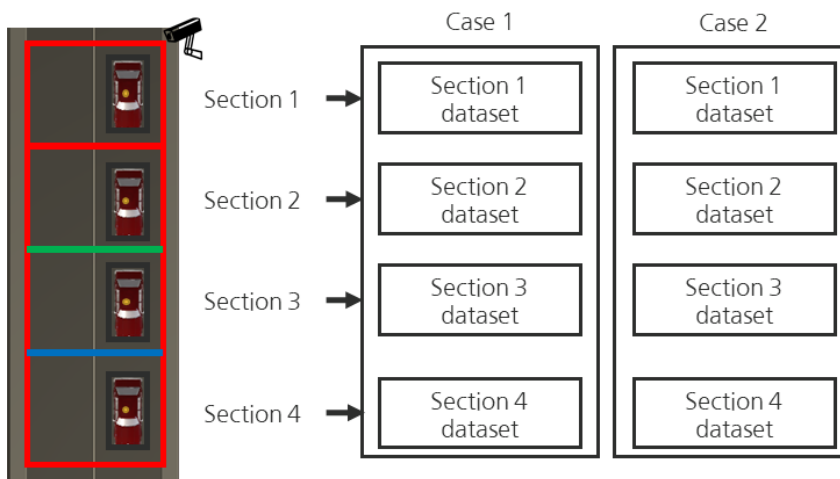


Fig. 9. Criteria for dividing sections in one image

이에 따라 Fig. 9와 같이 터널 CCTV로부터의 거리가 200 m 지점까지 관심영역을 설정한 다음 50 m 간격으로 4개의 구간으로 영역을 분할하여 각 구간마다 Case별 AP값 비교를 진행하였다. 각 구간별로 차량 객체를 구분하는 기준은 차량 경계박스의 y 최소값과 y 최대값의 평균을 기준으로 구분하였다.

본 논문은 같은 객체에 대하여 레이블링된 경계박스와 추론된 경계박스를 비교할 때, 두 경계박스간 겹치는 비율을 말하는 IoU (Intersection over Union)를 기준으로 참 또는 거짓을 판별하였으며(Padilla et al., 2021) 본 논문에서는 참과 거짓을 판별하는 하이퍼파라미터인 IoU의 임계값(Threshold)을 0.5로 설정하였다.

3.3.3 실험 조건

두 Case에 대한 딥러닝 데이터셋은 상기 설명한대로 동일한 영상에 대해 조건을 나눈 다음 레이블링을 진행하였으며 그 현황을 Table 1에서 확인할 수 있다. Table 1에서 두 Case는 모두 영상의 폭과 높이가 각각 646, 324로 동일하다. 딥러닝 데이터셋은 학습용 데이터셋(Train dataset)과 검증용 데이터셋(Test dataset)으로 나누었으며, 두 데이터셋에 대한 비중은 정지영상 개수 기준 각각 8:2의 비율로 나누었다. 본 논문에서 두 Case의 AP값 비교는 검증용 데이터셋에서만 비교를 진행할 것이므로, 구간 분할은 검증용 데이터셋만 진행하였다.

Table 1. Deep learning dataset status for each case

Experiment case	Dataset type	Experiment section	Number of images	Number of car objects
Case 1	Train	-	421	1,387
	Test	Section 1	58	111
		Section 2	55	87
		Section 3	63	83
		Section 4	70	92
		Total	106	373
Case 2	Train	-	421	1,387
	Test	Section 1	59	104
		Section 2	63	91
		Section 3	62	83
		Section 4	67	95
		Total	106	373

딥러닝 모델의 하이퍼파라미터의 경우, 학습율(Learning rate)는 0.0001, 정지영상에 대한 배치크기(Batch size)는 1, 학습회수(Epoch)는 100으로 설정하여 학습을 진행하였다. 합성곱 신경망층은 50층으로 구성된 ResNet (Residual network)을 사용하였다(He et al., 2016).

그리고 실험 대상 하드웨어 및 소프트웨어 조건은 AMD RYZEN 1800X, NVIDIA GTX 1080TI 11GB 사양을 가진 컴퓨터와 Python 3.7, Tensorflow 1.15.5 버전 소프트웨어로 딥러닝 모델의 학습을 진행하였다.

3.3.4 실험 결과

본 논문의 실험결과는 Table 2와 같이 4개의 구간으로 나누어진 검증용 데이터셋에 대한 AP값으로 비교 검토하였다. 추가적으로 보다 상세한 분석을 위해 Table 2에서는 각 Case마다 AP값 증감률을 기술하였다. AP값 증감률은 각 Section에서 Section 1과 비교하였을 때 AP값이 얼마나 감소하거나 증가하였는지를 백분율로 나타낸다.

Table 2. AP value result and the rate of change according to the inference by section of the deep learning model for each case in the test dataset

Experiment case	Category	Section 1	Section 2	Section 3	Section 4
Case 1	AP value	0.9737	0.9153	0.8169	0.8595
	Rate of change	0.00%	-6.00%	-16.10%	-11.73%
Case 2	AP value	0.9369	0.9457	0.8971	0.9647
	Rate of change	0.00%	0.94%	-4.25%	2.97%

먼저 Case 1에서는 Section 1의 AP가 0.9737으로 매우 높은 값을 보이고 있다. 이는 같은 Section에서 Case 2보다 0.0368 높은 수치 값인데, 원근현상에 의해 터널 CCTV와 가까운 지점에서는 객체가 매우 크게 보이므로 큰 객체들에 대한 딥러닝 모델의 객체인식 성능이 가장 높았다고 판단할 수 있다. 그런데 Section 2에서 Case 1의 AP값은 0.9153, Section 3에서는 0.8169, Section 4에서는 0.8595로 감소하였다. 이러한 경향을 증감률로 환산하여 분석해보면 터널 CCTV와 멀어질수록 AP가 크게 하락하는 경향을 확인할 수 있다. 즉 원본영상은 겹보기 속도 및 객체크기를 비교할 때와 마찬가지로 원근현상의 영향으로 인해 터널 CCTV와 먼 지점에서 딥러닝 모델의 객체인식 성능이 떨어지는 것으로 판단된다.

한편 Case 2에서 Section 1의 AP는 Case 1의 AP값보다 작정 정확도 수준에서 다소 낮지만, Section 2부터 Section 4까지 Case 2의 AP값은 같은 Section에서 Case 1의 AP값보다 각각 0.0304, 0.0802, 0.1052 큰 정확도 차이를 보였다. 증감률로 판단해보면 Case 2에서는 터널 CCTV와 멀어져도 AP가 감소하지 않고 비교적 일정한 정확도 성능을 유지하는 경향을 보인다. 이는 Case 2의 경우 역 원근변환을 통하여 먼 지점의 차량 객체의 크기가 가까운 지점에 있을 때 크기와 유사하게 유지되는데 기인한 것으로 판단된다.

이러한 결과들을 종합했을 때 Case 2와 같이 역 원근변환을 통해 다소 일그러진 객체의 형상이나 선명도가 떨어지는 특성 보다는 객체 형상의 일관성을 유지할 수 있다면 객체의 크기 유지가 객체인식 성능에 더욱 유리하게 작용한다는 것을 알 수 있다.

4. 결론

본 논문은 터널 CCTV에서 발생하는 심한 원근현상을 극복하기 위해 터널 영상유고시스템에서 역 원근변환 기법을 도입하였다. 이 기법을 통한 원근현상 완화를 확인하기 위해 동일한 관심영역을 설정한 다음 원본영상과

변환영상을 생성하였다. 본 논문의 실험은 먼저 정속으로 주행하는 차량에 대해 실제 위치에 따른 겉보기 속도와 크기를 비교함으로써, 변환영상에서 원근현상으로 인한 속도 및 크기의 변화에 대한 영향을 줄일 수 있다는 것을 확인하였다.

그 다음 원본영상과 변환영상에 대한 딥러닝 데이터셋을 각각 제작하여 동일한 조건으로 딥러닝 모델을 학습하였다. 그리고 딥러닝 모델의 객체인식 성능 분석은 일반적인 터널 내 CCTV 설치간격인 200 m 거리까지 거리별로 50 m씩 4개의 구간으로 나누어 구분하여 수행하였다. 그 결과 터널 CCTV와 최 인접한 구간에서는 원본영상이 유리할 수 있지만 멀어 질수록 변환영상을 사용하는 것이 확연히 우수한 객체인식 성능을 낼 수 있음을 알 수 있었다. 또한 변환영상을 사용한 경우는 50 m 이내 근접 구간에서 확보할 수 있는 수준의 객체인식 성능을 대상객체가 멀어져도 어느정도 객체인식 성능을 유지할 수 있음을 알 수 있었다. 이러한 결과들을 종합했을 때, 다음과 같은 결론을 얻을 수 있었다.

1. 터널 영상유고시스템을 구성할 때, 터널 CCTV 간격이 50 m 이내의 초 근접 구간에서는 원본영상을 이용해도 어느정도 객체인식 성능 확보가 가능하다. 따라서 대부분 원본영상을 사용한 기존의 영상유고시스템들도 가탐 거리를 제한한다면 큰 문제가 없을 것으로 판단된다. 즉 원본영상을 사용하는 터널 영상유고시스템을 사용할 시, 터널 CCTV 설치 간격을 조밀하게 추가 설치한다면 영상유고시스템의 운영상에 큰 문제가 없을 수 있다.
2. 하지만 터널 CCTV 간격이 50 m를 초과하는 현장에서는 변환영상을 사용하는 것이 유리하다. 이는 대부분의 터널 현장이 도로터널 방재시설 설치 및 관리지침에 의거하여 200 m 전후 간격으로 설치되어 있고, 원근효과의 영향을 최소화한 변환영상을 사용한다면, 영상유고시스템 운영을 위한 추가적인 CCTV 설치 없이도 기 설치된 CCTV 설비를 활용하여 영상유고시스템을 연결해 음영구간없이 자동 영상유고 관제를 진행할 수 있음을 의미한다.

본 논문에서는 터널 내 열악환경 조건을 감안하여 거리별 영상유고 성능을 이론적으로 고찰하기 위해 가상 터널 영상을 제작하여 실험에 활용하였으나, 다수 차량들의 앞뒤 겹쳐보임 등 원근현상과 연계된 보다 열악한 현장 영상조건에서 본 논문에서 제안한 변환영상 활용 효과가 더욱 배가될 수 있다. 이에 대해서는 실제 터널현장의 CCTV 영상을 통한 연구를 진행 후 차기 논문을 통해 추가 보고될 예정이다.

감사의 글

본 연구는 과학기술정보통신부 한국건설기술연구원 연구운영비지원(주요사업)사업으로 수행되었습니다(과제번호 20220124-001, 극한건설 환경 구현 인프라 및 TRL6 이상급 극한건설 핵심기술 개발).

저자 기여도

이규범은 연구 개념 및 설계, 원고 작성을 하였고, 신휴성은 원고 검토, 데이터 수집 및 데이터 분석을 하였다.

References

1. Bertozz, M., Broggi, A., Fascioli, A. (1998), "Stereo inverse perspective mapping: theory and applications", *Image and Vision Computing*, Vol. 16, No. 8, pp. 585-590.
2. Eggert, C., Brehm, S., Winschel, A., Zecha, D., Lienhart, R. (2017), "A closer look: small object detection in faster R-CNN", *Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME)*, Hong Kong, pp. 421-426.
3. Galeano, D.B., Devy, M., Boizard, J.L., Filali, W. (2011), "Real-time architecture on FPGA for obstacle detection using inverse perspective mapping", *Proceedings of the 2011 18th IEEE International Conference on Electronics, Circuits, and Systems*, Beirut, pp. 788-791.
4. Gyeonggi Province (2022), Road tunnel information status (open standard).
5. He, K., Zhang, X., Ren, S., Sun, J. (2016), "Deep residual learning for image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, pp. 770-778.
6. Juliani, A., Berges, V.P., Teng, E., Cohen, A., Harper, J., Elion, C., Goy, C., Gao, Y., Henry, H., Mattar, M., Lange, D. (2018), "Unity: A general platform for intelligent agents", *arXiv preprint*, arXiv:1809.02627.
7. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L. (1990), "Handwritten digit recognition with a back-propagation network", *Advances in Neural Information Processing Systems*, pp. 396-404.
8. LeCun, Y., Yoshua, B., Hinton, G. (2015), "Deep learning", *Nature*, Vol. 521, pp. 436-444.
9. Lee, E.S., Choi, W., Kum, D. (2019), "Bird's eye view localization of surrounding vehicles: longitudinal and lateral distance estimation with partial appearance", *Robotics and Autonomous Systems*, Vol. 112, pp. 178-189.
10. Lee, H.S., Jeong, S.H., Lee, J.W. (2011), "Real-time lane violation detection system using feature tracking", *The KIPS Transactions: Part B*, Vol. 18, No. 4, pp. 201-212.
11. Mallot, H.A., Bülthoff, H.H., Little, J.J., Bohrer, S. (1991), "Inverse perspective mapping simplifies optical flow computation and obstacle detection", *Biological Cybernetics*, Vol. 64, No. 3, pp. 177-185.
12. Min, Z., Ying, M., Dihua, S. (2019), "Tunnel pedestrian detection based on super resolution and convolutional neural network", *Proceedings of the 2019 Chinese Control and Decision Conference (CCDC)*, Nanchang, pp. 4635-4640.
13. MOLIT (2010), ITS project implementation guidelines, Ministry of Land, Infrastructure and Transport.
14. MOLIT (2021), Guideline of installation and management of disaster prevention facilities on road tunnels, Ministry of Land, Infrastructure and Transport.
15. Padilla, R., Passos, W.L., Dias, T.L., Netto, S.L., da Silva, E.A. (2021), "A comparative analysis of object detection metrics with a companion open-source toolkit", *Electronics*, Vol. 10, No. 3, 279.

16. Pflugfelder, R., Bischof, H., Dominguez, G.F., Nolle, M., Schwabach, H. (2005), "Influence of camera properties on image analysis in visual tunnel surveillance", Proceedings of the 2005 IEEE Intelligent Transportation Systems, Vienna, pp. 868-873.
17. Ren, S., He, K., Girshick, R., Sun, J. (2015), "Faster R-CNN: Towards real-time object detection with region proposal networks", Advances in Neural Information Processing Systems, Montréal.
18. Shin, H.S., Kim, D.G., Yim, M.J., Lee, K.B., Oh, Y.S. (2017a), "A preliminary study for development of an automatic incident detection system on CCTV in tunnels based on a machine learning algorithm", Journal of Korean Tunnelling and Underground Space Association, Vol. 19, No. 1, pp. 95-107.
19. Shin, H.S., Lee, K.B., Yim, M.J., Kim, D.G. (2017b), "Development of a deep-learning based tunnel incident detection system on CCTVs", Journal of Korean Tunnelling and Underground Space Association, Vol. 19, No. 6, pp. 915-936.
20. Tong, K., Wu, Y., Zhou, F. (2020), "Recent advances in small object detection based on deep learning: a review", Image and Vision Computing, Vol. 97, 103910.
21. Tsai, C.M., Hsieh, J.W., Shih, F.Y. (2016), "Motion-based vehicle detection in Hsuehshan Tunnel", Proceedings of the 2016 Eighth International Conference on Advanced Computational Intelligence (ICACI), Chiang Mai, pp. 385-389.
22. Zhu, M. (2004), Recall, precision and average precision, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo.
23. Zou, Z., Shi, Z., Guo, Y., Ye, J. (2019), "Object detection in 20 years: a survey", arXiv preprint, arXiv: 1905.05055.