

단안 깊이와 키포인트 기반 위치 추정을 활용한 경량 차량-보행자 위험 평가 파이프라인 연구

장현준, 유재천*

성균관대학교 전자전기컴퓨터공학과

10hour@skku.edu, *yoojc@skku.edu

A Study on the Lightweight Vehicle-Pedestrian Risk Assessment Pipeline using Monocular Depth and Keypoint-based Localization

Hyunjun Jang, Jae-Chern Yoo*

Dept. of Electrical and Computer Engineering, Sungkyunkwan University

요약

본 논문은 단일 RGB 이미지만으로 차량과 보행자 간의 충돌 위험을 실시간으로 판단하는 경량화된 위험 평가 시스템을 제안한다. YOLOv8 기반의 객체 및 자세 추정과 Depth Anything V2를 융합하여 고가의 3D 센서 없이도 실시간 처리가 가능한 저비용 ADAS 솔루션으로서의 유효성을 입증하였다.

I. 서론

본 논문에서는 단일 RGB 카메라 영상만으로 도로 상의 차량과 보행자 간의 3 차원 공간 관계를 정밀하게 분석하는 실시간 위험 평가 프레임워크를 제안한다.

도로 환경에서 보행자 안전은 자율주행 및 첨단 운전자 보조 시스템(ADAS)의 핵심 과제이다. 기존의 보행자 안전 평가 방식은 크게 두 가지로 분류된다. 첫째, 2D 이미지 상의 픽셀 좌표만을 활용하는 방식은 원근 왜곡(perspective distortion)으로 인해 실제 물리적 거리를 정확히 추정하기 어렵다. 둘째, LiDAR 나 스테레오 카메라를 활용한 3D 재구성 방식은 정확도가 높으나, 고비용의 센서가 필요하며 연산 복잡도가 높아 실시간 처리에 한계가 있다.

본 연구에서는 단일 RGB 카메라만을 사용하여 차량과 보행자 간의 상대적 위치 및 위험도를 실시간으로 평가하는 경량 시스템을 제안한다. 제안 시스템은 객체 검출(Object Detection), 자세 추정(Pose Estimation), 단안 깊이 추정(Monocular Depth Estimation), 차량 방향 추정(Vehicle Orientation Estimation)을 통합하여, 추가 센서 없이 Bird's Eye View(BEV) 기반의 공간적 위험도 분석을 수행한다.

II. 본론

제안 시스템의 전체 파이프라인은 [그림 1]과 같다. 입력 RGB 이미지로부터 차량 검출, 보행자 자세 추정, 깊이 추정을 병렬로 수행한 후, 좌표 변환을 통해 Bird's Eye View 표현으로 변환한다.

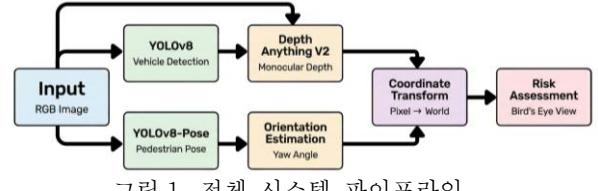


그림 1. 전체 시스템 파이프라인

차량 검출에는 YOLOv8n 모델을 사용하며, 보행자의 발목(ankle) 위치 추정에는 YOLOv8n-pose 모델을 적용한다. 발목 키포인트는 보행자가 지면과 접촉하는 위치를 나타내므로, 바운딩 박스 중심점보다 정확한 깊이 샘플링 지점을 제공한다. [그림 2(a)]는 검출 결과를 보여준다.

깊이 추정에는 Depth Anything V2 모델을 사용하여 입력 이미지로부터 dense depth map을 생성[그림 2(b)] 한다. 상대적 깊이 값 \hat{d} 는 다음과 같이 metric depth d 로 변환된다.

$$d = \frac{\alpha}{\hat{d} + \epsilon}$$

여기서 α 는 스케일 파라미터, ϵ 은 수치 안정 상수이다.

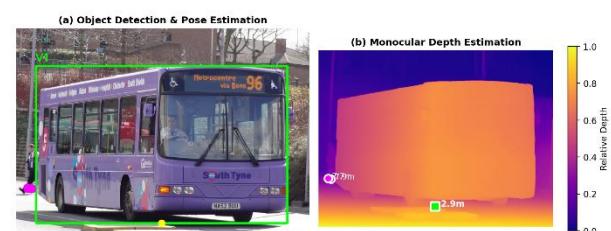


그림 2. 검출 및 깊이 추정 결과

픽셀 좌표 $\mathbf{p} = (u, v)^T$ 에서 월드 좌표 $\mathbf{P} = (X, Y, Z)^T$ 로의 변환은 핀홀 카메라 모델에 기반하여, 다음과 같이 정의된다:

$$X = \frac{d}{f_x}(u - c_x), \quad Y = d, \quad Z = \frac{d}{f_y}(v - c_y)$$

여기서 f_x, f_y 는 초점 거리, (c_x, c_y) 는 주점(principal point)이다.

차량의 진행 방향(Yaw angle)은 바운딩 박스의 종횡비(aspect ratio) $r = w/h$ 를 기반으로 추정한다:

$$\psi = \begin{cases} 0^\circ & \text{if } r < 0.8 \text{ (정면)} \\ \pm 90^\circ & \text{if } r > 2.0 \text{ (측면)} \\ \text{linear interpolation} & \text{otherwise} \end{cases}$$

Bird's Eye View [그림 3]에서 차량 위치 \mathbf{P}_v 와 보행자 위치 \mathbf{P}_p 간의 유clidean 거리는 다음과 같이 계산된다:

$$d_{vp} = |\mathbf{P}_v - \mathbf{P}_p|_2 = \sqrt{(X_v - X_p)^2 + (Y_v - Y_p)^2}$$

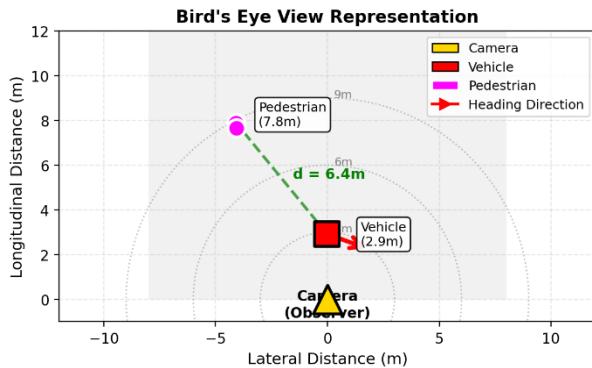


그림 3. Bird's Eye View 표현

또한, 차량 진행 방향 벡터 \vec{h} 와 차량-보행자 방향 벡터 \vec{d} 간의 각도 θ 를 통해 보행자가 차량의 전방/후방에 위치하는지 판단한다:

$$\theta = \arccos\left(\frac{\vec{h} \cdot \vec{d}}{|\vec{h}||\vec{d}|}\right)$$

$\theta > 90^\circ$ 인 경우 보행자는 차량 진행 방향의 후방에 위치하여 충돌 위험이 낮다고 판단한다.

III. 결 론

본 논문에서는 단일 RGB 이미지만으로 차량-보행자 간 공간적 관계를 분석하고 위험도를 평가하는 실시간 시스템을 제안하였다. 제안 방법은 고가의 3D 센서 없이도 Bird's Eye View 기반의 직관적인 위험도 시각화를 제공하며, 경량 모델 조합을 통해 실시간 처리가 가능하다.

[그림 3]의 실험 결과에서 차량(V1)은 카메라로부터 2.9m 거리에서 우측(Yaw 70°)으로 진행 중이며, 보행자는 7.9m 거리의 좌측 후방에 위치한다. 차량-

보행자 간 거리는 6.2m로, 보행자가 차량 진행 방향의 반대편에 위치하므로 충돌 위험이 낮음을 확인하였다.

향후 연구에서는 Optical Flow 기반의 속도 추정을 통합하여 충돌 예상 시간(Time-to-Collision, TTC)을 계산하고, 동적 위험도 평가로 확장할 예정이다. 또한, 경로 예측(Trajectory Prediction) 기법을 적용하여 미래 충돌 가능성을 사전에 예측하는 연구를 진행할 계획이다.

참 고 문 헌

- [1] Jocher, G., et al., "Ultralytics YOLO (Version 8.0.0) [Computer software]," 2023.
- [2] Yang, L., et al., "Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data," Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 10371–10381.
- [3] Cao, Z., et al., "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI), issued Jan. 1, 2021.