2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

Camera-LiDAR Jaywalking Detection in Traffic Surveillance System

1st Taek-Lim Kim
Dept. of Robot Control Eng
Chungbuk National University
Cheongju, Republic of Korea
taeglem@cbnu.ac.kr

2nd Byung Jin Jang
Dept. of Intelligent Systems and Robotics
Chungbuk National University
Cheongju, Republic of Korea
2023298004@cbnu.ac.kr

3rd Joo Yeon Yeon
Dept. of Intelligent Systems and Robotics
Chungbuk National University
Cheongju, Republic of Korea
yjyeon@cbnu.ac.kr

4th Tae-Hyeong Kim
Dept. of Research & Development
Korea Intelligent Automotive Parts Promotion Institute (KIAPI)
Daegu, Republic of Korea
thkim@kiapi.or.kr

5th Tae-Hyoung Park*
Dept. of Intelligent Systems and Robotics
Chungbuk National University
Cheongju, Republic of Korea
taehpark@cbnu.com

***Abstract*—Roadside sensors like cameras and LiDAR enhance pedestrian safety by providing comprehensive traffic data. While traditional traffic surveillance systems primarily focus on vehicle-related violations, such as signal violations and speeding, pedestrian jaywalking remains a significant cause of accidents. This paper presents a jaywalking detection method that fuses camera-based image segmentation with LiDAR ground segmentation to handle various conditions, including day and night. Our system addresses challenges such as poor lighting and vehicle occlusion by leveraging LiDAR's robustness in unlearned environments. Road segmentation is enhanced by combining camera outputs with LiDAR ground data, refining road boundary detection for more accurate road area analysis. By integrating road segmentation and object tracking, the system reduces false negatives and improves jaywalking detection. Experimental results from real-road data validate its effectiveness, showing significant potential to enhance traffic surveillance.**

***Index Terms*—Deep learning, Sensor Fusion, Intelligent Transportation Systems.**

I. Introduction

Jaywalking refers to a pedestrian crossing the road outside the designated crosswalk or disregarding traffic light signals. Such behavior can lead to severe accidents and result in numerous casualties. Therefore, it is crucial to detect objects and road areas to identify jaywalking, regardless of the time of day or environmental conditions. Previous studies [1], [2] on jaywalking have mainly concentrated on the relationship between vehicles and pedestrians or have examined vehicle violations by analyzing object detection, tracking, and behavior [3], [4]. In detecting jaywalking, the road area in the image is defined, and jaywalking is assessed based on this area and the location of detection when a person is identified [5]. Furthermore, deep learning-based approaches [6] identify anomalies in specific image frames and detect jaywalking based on object detection results in the anomalous frame.

However, existing camera-based jaywalking detection systems face significant challenges in low-light conditions, such as nighttime environments, or situations with intense light, like vehicle headlights, which can cause performance degradation [7]–[9]. Moreover, the systems struggle with vehicle occlusion, making it difficult to detect pedestrians crossing in front of or behind vehicles. To overcome these limitations, our approach implements sensor fusion, leveraging the complementary strengths of both camera and LiDAR data. The fusion of LiDAR, less affected by lighting conditions, with camera-based image segmentation, provides robust road and object detection performance in various environmental conditions, including day and night scenarios.

Our system performs object and road segmentation through a camera-based network and LiDAR ground segmentation [10]. The road segmentation is refined by merging the results from both sensors, allowing for more accurate edge detection and road area analysis. Additionally, the system can detect jaywalking violations based on defined road areas and object positions by fusing crosswalk and object tracking data. The crosswalk detection utilizes Hough transform [11] and MSER [12], further improving accuracy. This robust approach significantly reduces the false negative rate by considering the road area and improves jaywalking detection in challenging scenarios such as vehicle occlusion.

Our system has been thoroughly tested and validated with real-road data, demonstrating robust performance in various environments and situations. The results indicate that this approach can provide a reliable and adaptable solution for jaywalking detection, contributing to pedestrian safety and improving overall traffic surveillance systems.

This work was supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(IITP-2025-RS-2020-II201462).

*Corresponding author.

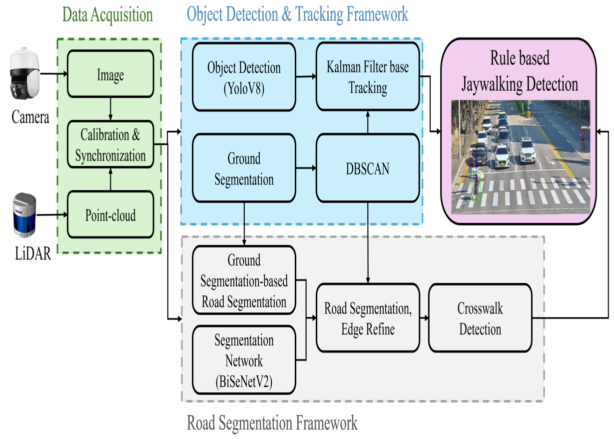
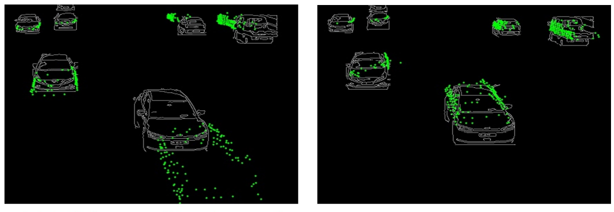


Fig. 1. The overall structure of the jaywalking detection system. The system is divided into two main parts: Road Segmentation Framework and Object Detection and Tracking Framework.



(a) Without calibration and synchronization

(b) Applying calibration and synchronization

(c) Synchronized data

Fig. 2. Calibration and Synchronization: (a) Misaligned data without calibration. (b) Data after calibration and synchronization. (c) Fully synchronized LiDAR points projected onto the image.

II. Jaywalking Detection System

Fig. 1 shows an overview of the jaywalking detection system, which integrates data from both camera and LiDAR sensors. Initially, calibration and synchronization are performed to ensure that data from both sensors are properly aligned. Calibration is done using a targetless approach, making it suitable for real-world traffic environments where artificial landmarks may not always be practical. Synchronization ensures temporal and spatial consistency, which is crucial for effective fusion. All results presented in this paper are based on fully calibrated and synchronized data (see Fig. 2).

Once calibrated, the system processes data through two main frameworks: road segmentation and object detection and tracking. The segmentation framework utilizes BiSeNetV2 [8] for image segmentation, which is effective under both day and night conditions. Additionally, LiDAR data is used to create a ground mask, which is then fused with the segmented image to accurately define road boundaries.

The object detection and tracking framework is responsible for identifying and tracking pedestrians and other relevant objects. Camera-based detection uses YOLOv8 for classification, ensuring reliable detection under various lighting conditions. For LiDAR-based detection, DBSCAN [13] is applied to cluster LiDAR points after ground removal, effectively identifying pedestrians in 3D space. The fusion of detection results from both sensors ensures robustness; if a pedestrian is detected by the camera, the LiDAR depth information confirms and enhances the detection. Furthermore, in scenarios where the camera fails to detect an object due to occlusion or poor lighting, the LiDAR-based tracking continues, ensuring uninterrupted tracking of pedestrians. DeepSORT [14] is used to track objects detected by the camera, while LiDAR data further improves tracking by providing reliable distance measurements.

Crosswalk detection is another key component of the system. This is achieved by extracting lines through edge detection and using the Hough transform to identify crosswalks. The crosswalk regions are defined based on these detections, and they are used as reference zones to determine jaywalking events.

A. *Road Segmentation, Crosswalk Detection*

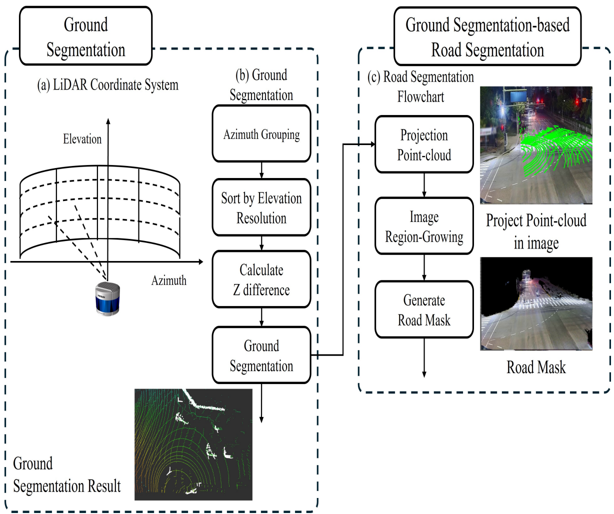


Fig. 3. LiDAR coordinate system description and ground segmentation. Road area segmentation flowchart: (a) LiDAR coordinate system: depiction of the LiDAR's elevation and azimuth angles. (b) Ground segmentation: steps include azimuth grouping, elevation sorting, and z-axis difference calculation. (c) Road area segmentation: projecting the ground's point cloud onto the region-growing image.

The result of the image segmentation network is called M^c. Ground segmentation is shown in Fig. 3, and the LiDAR point cloud can be expressed in azimuth and elevation coordinates. We grouped the LiDAR points by azimuth, sorted the groups by elevation, and then detected the ground as a threshold value through the z-axis difference from the previous point. Fig. 3. (b) is the result of classifying the ground as a color and white as an edge. Fig. 3. (c) results from projecting the

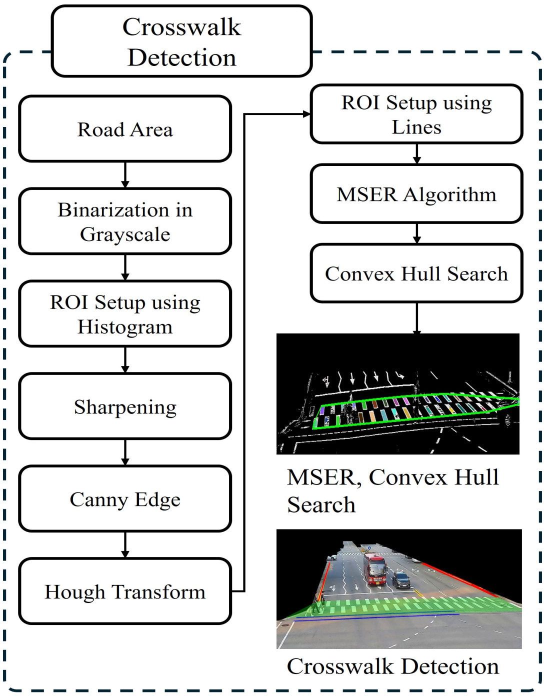


Fig. 4. Crosswalk detection flowchart: Crosswalk detection involves road area segmentation, Hough transform, MSER, and convex hull search.

ground. A mask is generated by applying region growing to the projected image [15]. The road mask generated from the ground segmentation results is called M^l, and the road segmentation is completed by merging M^c and M^l.

Fig. 4 illustrates the flowchart for crosswalk detection. The process starts by converting the road image to grayscale and setting up the Region of Interest (ROI) using a histogram analysis to focus on potential crosswalk areas. After defining the ROI, edge detection and the Hough transformation are applied to highlight white areas, particularly crosswalk markings.

Once the white lines are identified through binarization, the system selects line segments with slopes as close to zero as possible, as these likely represent horizontal crosswalk markings. Using the MSER (Maximally Stable Extremal Regions) algorithm, the detected regions are further refined, and the convex hull algorithm is applied to connect the segmented regions, allowing for more precise crosswalk detection and highlighting.

B. Rule of Jaywalking Detection

Fig. 5. presents a flowchart outlining the process of determining whether a pedestrian is jaywalking. The flowchart consists of two steps. First step: A person standing near the lanes within the road area but not on the road is identified as a potential jaywalker, avoiding the classification of all cases where a pedestrian stands in a road area as jaywalking, which previously led to many false positives. So, we set up a warning zone to represent a potential hazard. The result is a boundary on the road mask and a crosswalk boundary. Next, Pedestrians who are not in a warning zone are judged to be jaywalking if they are not projected onto the crosswalk after determining whether they are on the crosswalk.

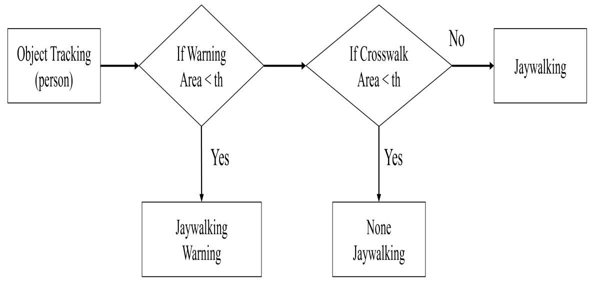


Fig. 5. Jaywalking flowchart: Jaywalking is determined based on the location of the warning and crosswalk areas.

This method offers significant benefits for pedestrian safety. It involves using a network to detect objects from cameras and measure distances using LiDAR. Unlike the existing method, which detects objects in the road area using the measured distance from the camera and considers it a violation when a person walks on the road instead of the crosswalk, our method accurately classifies crosswalk violations as warnings rather than pedestrians near the roadside. This approach improves accuracy and reduces the number of false positives, ultimately enhancing overall pedestrian safety measures. Road area segmentation is necessary to implement this approach, utilizing the road area described in the previous section. Furthermore, the road area image is used to detect crosswalks.

III. Experiments

A. Experiment Setup

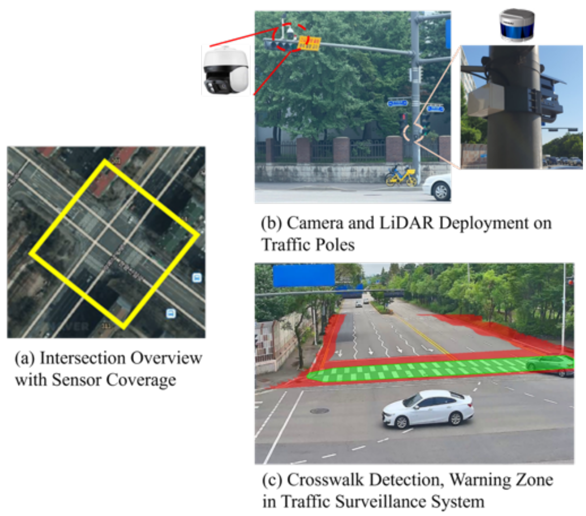


Fig. 6. Experimental environment: (a) Aerial view of the experimental area with sensor coverage, provided by Naver Map. (b) Deployment of camera and LiDAR sensors on traffic poles. (c) Crosswalk detection and warning zone configuration in the traffic surveillance system.

Fig. 6. shows data collected on traffic violations on real roads in Siheung City, South Korea, as an experimental environment. Fig. 6. (a) is an aerial photo provided by the

Naver Map application. The camera is attached to the traffic light in the intersection area, and the LiDAR is attached to the roadside, as in Fig. 6. (b). Fig. 6. (c) is set using the crosswalk and road area detected above as a crosswalk and warning zone. The green area is the crosswalk area, and the red area is the edge of the crosswalk and road mask, which is set as a warning zone to improve crosswalk violation detection.

The experimental environment was conducted at an intersection in Korea, and four CCTV cameras and two Hesai 40ch LiDARs were used. Road area detection evaluation: 128 images, consisting of 80 daytime and 48 nighttime images, were used for segmentation evaluation. The camera and LiDAR data are synchronized. The GPU used was RTX 3080, the detection network was Yolov8, and the segmentation network was BiSeNetv2. All network training was done solely with open dataset kitti and cityscape. Jaywalking detection evaluation was conducted on 1200 synchronized camera data and LiDAR data.

The threshold value for jaywalking detection was determined empirically through extensive testing with real-world data. This iterative process aimed to minimize both false positives (e.g., pedestrians incorrectly classified as jaywalking) and false negatives (e.g., missed detections of actual jaywalkers). By adjusting the threshold based on the system's performance under different lighting and traffic conditions, we were able to achieve a balanced level of sensitivity and specificity, ensuring reliable detection performance. Future work could further investigate the impact of varying this threshold to optimize performance metrics like precision, recall, and F1 score in different environments.

B. Experiments on Road Segmentation

TABLE I
ROAD SEGMENTATION PERFORMANCE: COMPARISON OF MEAN INTERSECTION OVER UNION (MIOU) FOR ROAD SEGMENTATION USING BISENETV2, UNDER BOTH DAY AND NIGHT CONDITIONS.

Modal	Network	Ground	mIoU-Day	mIoU-Night
Camera	BiSeNetv2	X	0.89	0.82
Camera+LiDAR	BiSeNetv2	O	**0.91**	**0.86**

TABLE II
OBJECT SEGMENTATION PERFORMANCE: COMPARISON OF MEAN INTERSECTION OVER UNION (MIOU) FOR OBJECT SEGMENTATION USING BISENETV2, UNDER BOTH DAY AND NIGHT CONDITIONS.

Modal	Network	Edge Refine	Accuracy-Day	Accuracy-Night
Camera	BiSeNetv2	X	0.30	0.35
Camera+LiDAR	BiSeNetv2	O	**0.38**	**0.39**

Table I shows the results of road segmentation. BiSeNet was trained using the cityscape dataset and evaluated using Siheung's data. Ground is a proposed method that complements road area segmentation by detecting the ground using LiDAR and region-growing methods. We achieved performance improvements of about 2 percent during the day and 5 percent at night.

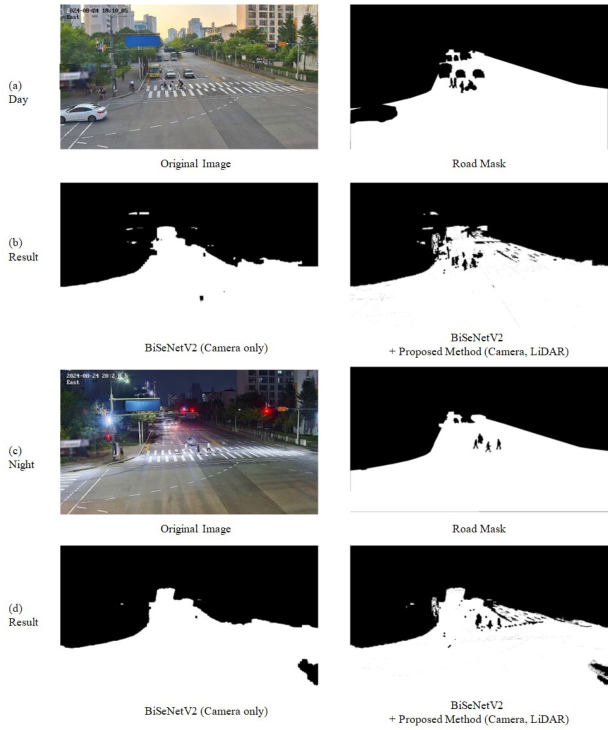


Fig. 7. Results of comparative experiments on detection of road areas: (a,b) Day segmentation results comparing BiSeNetV2 (Camera only) with the proposed method (Camera + LiDAR). (c,d) Night segmentation results comparing BiSeNetV2 (Camera only) with the proposed method (Camera + LiDAR).

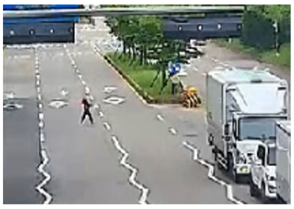
The proposed method restored some of the road areas that were not detected in Fig. 7 (b) and (d). Although Bisenet generally detected the road area effectively, it exhibited limitations in accurately segmenting objects. Table II highlights its difficulty in achieving precise object segmentation, and this limitation is further illustrated in Fig. 7.

C. Experiments on Jaywalking Detection

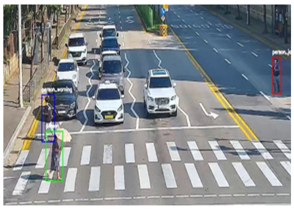
TABLE III
THE JAYWALKING DATASET SHOWS THE DISTRIBUTION OF FRAMES ACROSS DIFFERENT EPISODES DAY AND NIGHT.

Episode	Day	Night
None-jaywalking	2109	605
(E1) A jaywalker outside of the warning zone	366	180
(E2) A jaywalker standing in the warning zone	726	164
(E3) A jaywalker disappears behind a vehicle	30	96
Total	3231	1045

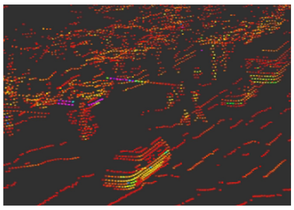
Fig. 8 illustrates different scenarios of crosswalk violations. Episode 1 involves a pedestrian crossing in an area that is neither a designated crosswalk nor a warning zone but rather a section where crossing is not permitted. Episode 2 involves entering the road area slightly away from the crosswalk. Episode 3 represents an occlusion jaywalking by a vehicle, which is difficult to detect in the image. Table III presents data on individuals and pedestrians crossing the street, with data collected during both day and night.



(a) E1 : Jaywalker Outside of the Warning Zone



(b) E2 : Jaywalker Standing in the Warning Zone.



(c) E3 : Jaywalker Disappears Behind a Vehicle (image, LiDAR)

Fig. 8. Jaywalking case study. (a) E1: A jaywalker outside of the warning zone. (b) E2: A jaywalker standing in the warning zone, potentially posing a hazard. (c) E3: Jaywalker disappears behind a vehicle, challenging for visual detection. left : caemra, right: LiDAR.

TABLE IV
JAYWALKING EPISODE LEVEL EVALUATION.
P : PRECISION, R : RECALL, F1 : F1 MEASURE

Modal	Episode	Total Frames	P	R	F1
Camera	Normal	2714	0.99	0.94	0.96
	E1	546	**0.95**	0.76	0.84
	E2	790	**0.99**	0.98	0.98
	E3	126	**1.0**	0.30	0.46
Camera+LiDAR	Normal	2714	**0.99**	**0.99**	**0.99**
	E1	546	0.94	**0.79**	**0.86**
	E2	790	0.97	**0.99**	**0.99**
	E3	126	0.99	**0.92**	**0.95**

Table IV shows the results of the episode-by-episode evaluation. Overall, the proposed method has improved performance, and in particular, the method that integrates LiDAR shows a significant difference in the E3 environment where occlusion occurs. Although the proposed method has slightly lower precision, it has significantly improved recall, demonstrating that LiDAR effectively compensates for missed detections when the camera fails to detect objects. Fig. 9. illustrates the detection and projection results of LiDAR clustered data in occlusion scenarios. The occlusion begins in Fig. 9 (a). In (b), the camera fails to detect the object, but the LiDAR point cloud is successfully projected. As shown in Fig. 9 (c), the proposed fusion method continues tracking even during the occlusion, whereas the camera-only method fails to track the object, leading to missed detection and lower recall. (e) shows the start of occlusion at night, while (f) presents the corresponding LiDAR data. (g) illustrates the performance of the proposed method, and (h) shows the camera-only method. After the occlusion ends, we can observe a noticeable difference in tracking performance between the two methods.

Table V shows the performance difference between the camera-only and LiDAR fusion methods during the day and night. The fusion of LiDAR during both day and night significantly improved overall performance, with recall increasing by 15 percent, particularly at night. Fig. 10 shows the detection results for both day and night, where the white background represents detection results, and the gray background indicates tracking results.

TABLE V
JAYWALKING DETECTION EVALUATION.
P : PRECISION, R : RECALL, F1 : F1 MEASURE

Modal	Day			Night		
	P	R	F1	P	R	F1
Camera [5]	0.99	0.93	0.96	0.97	0.82	0.89
Camera+LiDAR	**0.99**	**0.97**	**0.98**	**0.98**	**0.97**	**0.97**

IV. DISCUSSION AND CONCLUSION

The proposed jaywalking detection system leverages sensor fusion between camera and LiDAR, which is particularly advantageous in occlusion scenarios. While the strategic placement of the sensors helps expand coverage, the key benefit lies in the fusion of complementary data from both sensors, enhancing robustness beyond what placement alone could achieve.

LiDAR provides reliable depth information and maintains performance in poor lighting, complementing the camera's ability to classify objects effectively under optimal lighting conditions. Even if an additional camera were placed at the LiDAR's location, issues like nighttime occlusion would persist. Fusing both sensors ensures continuous detection, even when camera data is compromised, demonstrating the strength of sensor fusion over placement alone.

Although road segmentation could theoretically be performed under ideal conditions, real-world traffic environments are dynamic, with changing conditions throughout the day. Cars may be parked, obstacles may appear, and modifications may occur, making real-time segmentation essential for maintaining accuracy. By updating segmentation continuously, the system adapts effectively to changing road environments.

This study highlights the value of sensor fusion in jaywalking detection, particularly in minimizing false negatives. LiDAR improves segmentation, road detection, and tracking, especially during occlusions. Future work will focus on validating performance under adverse weather and improving real-time processing capabilities. Despite these challenges, the proposed system offers a solid foundation for enhancing traffic surveillance and pedestrian safety.

ACKNOWLEDGMENT

This work was supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(IITP-2025-RS-2020-II201462)

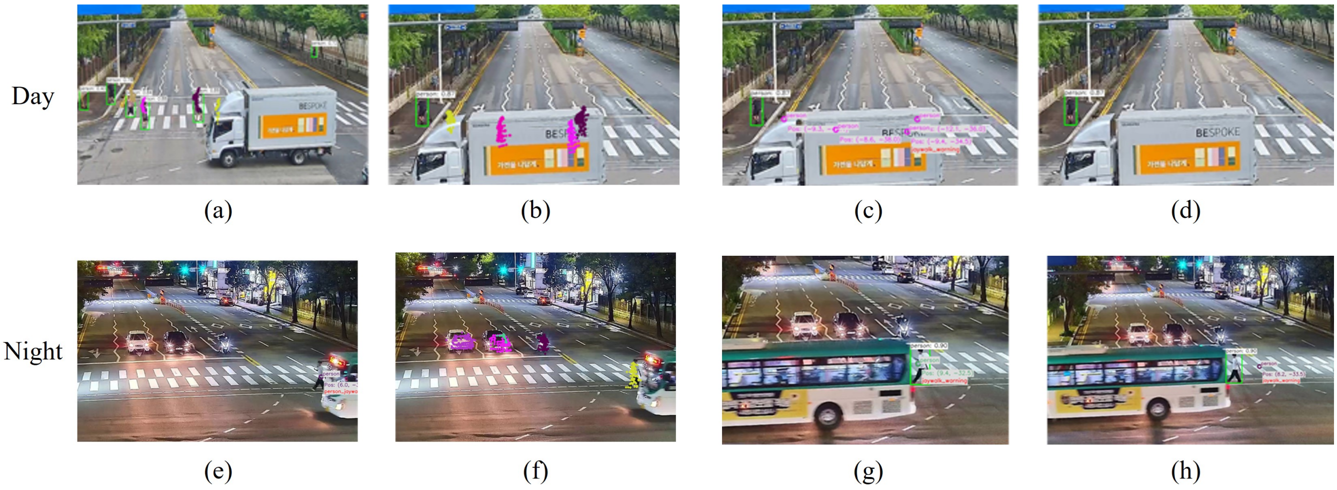
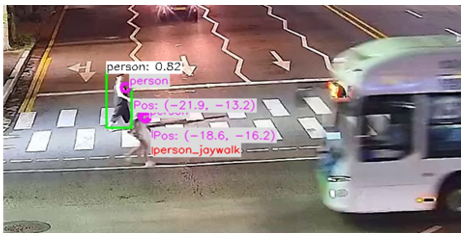


Fig. 9. Comparison of camera-only vs. camera+LiDAR for jaywalking detection. (a) Occlusion starts during the day with partial pedestrian detection. (b) LiDAR data projected during the peak occlusion. (c) The proposed method combining camera and LiDAR data. (d) Camera-only method during the day. (e) Occlusion starts at night. (f) LiDAR data during the start of occlusion. (g) The proposed method at night using both camera and LiDAR. (h) Camera-only method at night.

(a) Jaywalking Detection in Day



(b) Jaywalking Detection in Night

Fig. 10. Jaywalking detection performance. (a) Jaywalking detection during day, where pedestrians are accurately detected and labeled, including warning flags for those outside crosswalk boundaries. (b) Jaywalking detection at night, demonstrating the system's robustness under low-light conditions, with clear identification of pedestrians and jaywalking alerts.

References

[1] Y. Zheng, T. Chase, L. Elefteriadou, B. Schroeder, and V. P. Sisiopiku, "Modeling Vehicle–Pedestrian Interactions Outside of Crosswalks," *Simulation Modelling Practice and Theory*, vol. 59, pp. 89–101, 2015.

[2] B. R. Kadali and P. Vedagiri, "Proactive Pedestrian Safety Evaluation at Unprotected Mid-Block Crosswalk Locations Under Mixed Traffic Conditions," *Safety Science*, vol. 89, pp. 94–105, 2016.

[3] S. Sivaraman and M. M. Trivedi, "Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1773–1795, 2013.

[4] M. A. H. Anik, M. Hossain, and M. A. Habib, "Investigation of Pedestrian Jaywalking Behaviour at Mid-Block Locations Using Artificial Neural Networks," *Safety Science*, vol. 144, p. 105448, 2021.

[5] S. Mostafi, W. Zhao, S. Sukreep, K. Elgazzar, and A. Azim, "Real-time jaywalking detection and notification system using deep learning and multi-object tracking," in *Proc. 2022 IEEE Global Communications Conference (GLOBECOM)*, pp. 1164–1168, 2022.

[6] A. Pramanik, S. Sarkar, and J. Maiti, "Jaywalking detection and localization in street scene videos using fine-tuned convolutional neural networks," *Accident Analysis & Prevention*, vol. 154, p. 106019, 2021.

[7] D. Wu, L. Chen, and Y. Zhu, "YOLOP: You only look once for panoptic driving perception," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 3203–3213, 2021.

[8] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "BiSeNet V2: Bilateral Network With Guided Aggregation for Real-Time Semantic Segmentation," *International Journal of Computer Vision*, vol. 129, no. 11, pp. 3051–3068, 2021.

[9] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 420–430, 2023.

[10] W. Zhang, "Lidar-Based Road and Road-Edge Detection," in *2010 IEEE Intelligent Vehicles Symposium*, pp. 845–848, 2010.

[11] H. Liu and M. Lu, "A crosswalk stripe detection model based on gradient similarity tags," in *Proc. 2022 7th Int. Conf. Image, Vis. Comput. (ICIVC)*, pp. 114–122, 2022.

[12] Y. Zhai, G. Cui, Q. Gu, and L. Kong, "Crosswalk detection based on MSER and ERANSAC," in *Proc. 2015 IEEE Intell. Transp. Syst. Conf. (ITSC)*, pp. 2770–2775, 2015.

[13] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD)*, vol. 96, no. 34, pp. 226–231, 1996.

[14] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. 2017 IEEE Int. Conf. Image Process. (ICIP)*, pp. 3645–3649, 2017.

[15] Y.-L. Chang and X. Li, "Adaptive Image Region-Growing," *IEEE Transactions on Image Processing*, vol. 3, no. 6, pp. 868–872, 1994.