Driving Assistance: Pedestrians and Bicycles Accident Risk Estimation using Onboard Front Camera

Stephen Karungaru, Ryosuke Tsuji & Kenji Terada
Tokushima University
(2-1 Minami Josanjima Tokushima, Japan, karungaru@tokushima-u.ac.jp)

Abstract In this study, we propose a collision detection system by detecting and estimating the risk posed by pedestrians and bicycles in the images captured by a monocular onboard camera. In the proposed intrusion system. after initial detection, the pedestrians and bicycles are tracked to obtain their location, the direction of movement, and posture information using lane detection information, velocity calculation and, pose estimation respectively. Finally, this information is evaluated using fuzzy rules to estimate the risk the pedestrian and/or bicycle poses. The results are transmitted to the driver using voice and sound. We tested the system using 89 video scenes and achieved recall and precision accuracies of 0.94 and 0.87 respectively.

Keywords *Pedestrian, bicycle detection, Driver assistance, Danger posed estimation*

1 Introduction

Automobiles have become a life necessity for many people. However, many avoidable accidents still occur due to driver inattentiveness or sometimes rule ignoring pedestrians or bicycles. According to a survey by the Japan National Police Agency's Traffic Bureau, about 3,000 people lose their lives every year [1]. Active protection technology systems assist in dangerous driving conditions such as driver fatigue, distraction, blind areas, and pedestrian intrusion. Driver Fatigue Monitoring System (DFM), Blind Spot Detection (BSD), and Pedestrian Collision Warning (PCW) are the most important Advanced Driver Assistance Systems (ADAS) for the protection of pedestrians.

Pedestrian detection using onboard cameras is one of the most important topics in computer vision, with applications in areas such as advanced driver assistance systems, surveillance, safety systems, and advanced robotics. Many researchers have studied pedestrian detection using onboard cameras in recent years. Unlike the general pedestrian detection study, the study of pedestrian detection using the onboard camera is the detect pedestrians in a dynamic background from a moving camera. Existing studies can be classified according to the method they conducted and can be divided into the following categories:

Holistic detection: Holistic detectors are trained to detect pedestrians in images by scanning the whole frame. This kind of approach can reliably detect humans in a static image without motion information. In this method, related research works use different features to detect pedestrians, for example, [1] employed global features such as edge template, while [2] used local features like the histogram of oriented gradients descriptors. The drawback of this approach is that the performance can be easily affected by background clutter and occlusions. Even so, there are many research works related to the pedestrian detection system by modification or extension of this approach. [3], which used optical flow and Histogram of Oriented Gradients, is the most notable of this approach.

Part-based detection*:* Pedestrian detection with part-based approaches consists of collections of pedestrian parts. First, part hypotheses are derived by learning local features, like edgelet [4] and orientation features [5]. These part hypotheses are then combined to form the best assembly of pedestrian hypotheses. This approach is effective, but part detection is a challenging process.

Motion-based detection: In the research on pedestrian detection using onboard cameras, motion-based detection [6] is not effective due to conditions such as fixed camera, stationary lighting, etc.

Detection using multiple cameras: In this approach, the detector [7] produces a Probability Occupancy Map, which provides an estimation of the probability of each grid cell being occupied by a pedestrian. This kind of approach is mostly addressed in a surveillance system because this method integrates multiple calibrated cameras for detecting multiple pedestrians. Even so, some studies have used this method for the study of pedestrian detection.

Deep learning approach: Since Girshick et al. [8] proposed RCNN in 2014, the task of pedestrian detection has officially entered the deep learning stage. In general, detection methods based on deep learning mainly consist of two categories. One is a two-stage processing method (RCNN 2014 [8] Mask RCNN 2017 [9] Fast RCNN [10] and Faster RCNN [11]). Firstly, regional suggestion boxes for a possible object are generated, and further predictions are then made on these suggestion boxes. The other is a one-stage processing method (YOLOv1 [12], YOLOv2 [13], YOLOv3 [14], SSD [15], RetinaNet [16], DIOU [17], YOLOv4 [18] and YOLOv5), which directly returns the object area on the feature map and gives the final prediction result.

Conclusion: Although pedestrian detection technology has made great progress from the original traditional machine learning to the deep neural network, there is still a huge gap with human vision [19]. Regarding the application of pedestrian detection using deep learning, it is a necessity to ensure its real-time performance. Additionally, it is necessary to lighten the model, because it is difficult to implement in a practical driving environment. Therefore, shallow learning methods such as holistic detection versions are still the key method of real-time pedestrian detection using onboard camera systems.

Therefore, to reduce the number of fatalities caused by accidents, based on such research, ADAS systems in vehicles equipped with systems that support safe driving using advanced technologies and drive recorders equipment have been developed [21]. However, it is difficult to adopt because it is difficult to retrofit and is expensive.

Although pedestrian detection technology has made great progress from the original traditional machine learning to the deep neural network, there is still a huge gap with human vision [19]. Regarding the application of pedestrian detection using deep learning, it is a necessity to ensure its real-time performance. Additionally, it is necessary to lighten the model, because it is difficult to implement in a practical driving environment. Currently, functions to warn of approaching vehicles, lane out, speeding, and surrounding vehicles are common. However, there is no function to warn of dangerous pedestrians or bicycles.

As shown in the related works above, methods for pedestrians are numerous and great results have already been achieved. However, accurate detections alone might not offer a practical solution. Our work builds on these works by introducing fuzzy rules that determine the "intentions" of the pedestrian and warns the driver of the risk. Fuzzy rules use membership functions that can be defined as a technique to solve practical problems by experience rather than knowledge. Therefore, the idea is to put meaning to the detections and provide a practical solution to the driver. These "intention" includes "about to cross", "towards the car", etc. The fuzzy rules are created based on a combination of pedestrian, bicycle, lane, zebra, and person information.

Therefore, in this study, we propose a system that detects pedestrians' and bicycles' "intentions" using only a monocular onboard camera. The system obtains information such as their position, and direction of travel, and uses it to estimate the danger posed.

The method proposed in this work has one limitation: the use of one camera. In most recent works, 2 frontal cameras or one camera and a distance sensor are prevalent. The former is referred to as scene depth estimation. Data-driven approaches learn the depth using a supervised depth regressor [27] or unsupervised disparity estimator [28, 29]. The latter uses Radars (Radio Detection and Ranging) and Lidars (Light Detection and Ranging) for distance estimation. Our work uses one camera, the frame rate, and the speed of the car to estimate the distance to the detected object. While our distance estimation method may not be as accurate, it does not affect the effectiveness of the proposed method which is to estimate the risk pedestrians/bicycles pose to drivers.

The rest of this paper is organized as follows. Section 2 introduces and describes the proposed method in detail especially lane detection, person information acquisition, and risk assessment, section 3 presents the experiments conducted and the results, and finally, the conclusions and future works are presented.

2 Proposed Method

The objective of the proposed method is to determine the danger posed by pedestrians and bicycles based on fuzzy rules that are decided based on a combination of the detection results of pedestrian, bicycle, lane, Zebra crossing, and body movement determination. The details are provided in the sections below.

The flow chart below shows the proposed method. After the image is captured, brightness compensation is performed. The current lane and the presence of zebra crossing are continuously detected. If a pedestrian or a bicycle is also detected, the person's information (facing direction, position, direction of movement) is determined. Finally, based on these factors and the speed of the car, the risk posed is estimated and if high, the driver is warned of the danger.

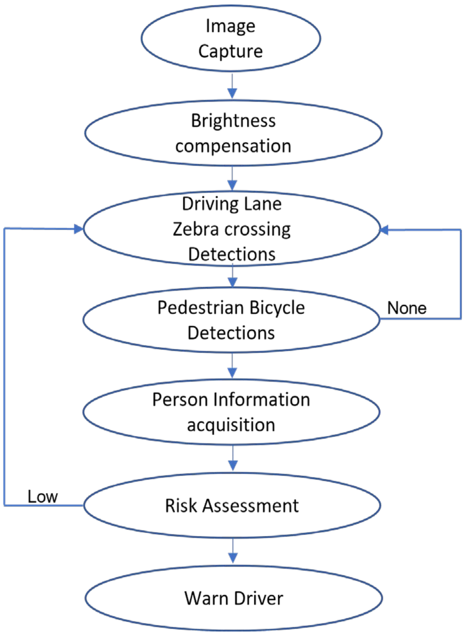


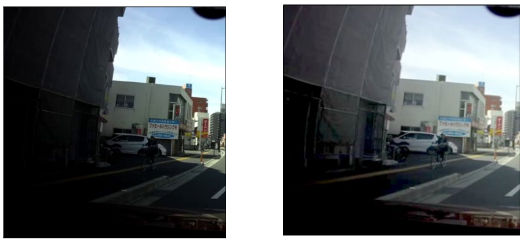
Image Capture
Brightness compensation
Driving Lane Zebra crossing Detections
Pedestrian Bicycle Detections
None
Person Information acquisition
Low
Risk Assessment
Warn Driver

Figure 1 : Process flowchart

2.1 Brightness Compensation

To ensure a stable image capture environment by the onboard camera, gamma correction is applied to adjust the contrast and suppress the influence of shadows. The conversion formula for gamma correction is shown in Equation (1), where X is the pixel value before gamma correction, Y is the pixel value after gamma correction, and γ is the gamma correction value. The result of gamma correction is shown in Figure 2.

$$Y = 255(\frac{X}{255})^{\frac{1}{\gamma}} \quad \text{......} (1)$$



(a) Before correction (b) After correction

Figure 2 : Brightness compensation

2.1 Lane Detection

The Canny method is used for edge detection and the Hough transform for straight line detection to detect lanes in images acquired from a camera. However, since processing the entire image is computationally expensive, by noting that the lane appears only at the bottom of the captured images, lane processing is only confined to this region. After edge detection in the regions, straight-line detection is performed using the Hough transform. Based on our camera position, the angle of the lines detected is limited to 40 to 60 degrees on the left and 120 to 140 degrees on the right. Among the lines detected under these conditions, the two innermost lines are extracted as lanes. However, the detection may fail when there is little white space in the dashed area or when the lane is blurred. If the detection fails, interpolation is performed using the straight-line detected immediately before. A sample result of lane detection is shown in Fig. 3.

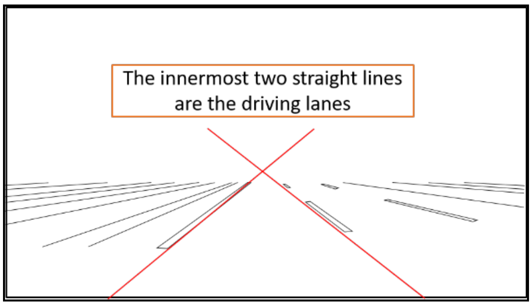


Figure 3 : Lane Detection

2.3 Zebra Crossing Detection

The presence or absence of a crosswalk is very important in estimating a person's crossing intention. Initially, we prepare several horizontal scan lines on the travel lane. The scanned lines are binalized based on the threshold value obtained by the discriminant analysis method and define the change points as the areas that alternate between black to white or white to black. Finally, a crosswalk is detected by using the variation, distribution, and the number of intervals between the change points, Fig. 4.

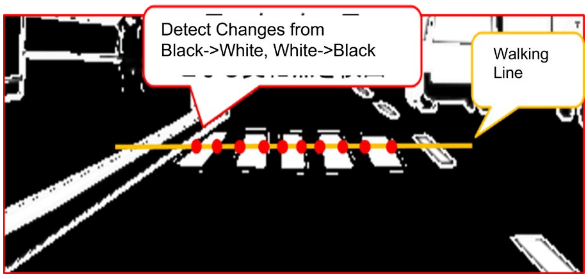


Figure 4 : Pedestrian Crossing Detection

2.4 Pedestrians and bicycles

2.4.1 Detection: The Single Shot MultiBox Detector (SSD) [22] is used for object detection. SSD is a kind of deep learning model that can detect the position of objects in an image and classify them simultaneously and fast. Since the target is a video taken by an in-vehicle camera, SSD was selected since it is effective for fast object classification and detection.

In SSD, the target image is resized into a square of a certain size and input to the network, which uses the network structure of the VGG16 model as its base [23], Fig. 5. The first feature map is the middle layer of the base network, and multiple feature maps of different sizes are created by further convolution. We classify and search for objects on each of the obtained feature maps. Each feature map is searched with a bounding box of a predefined size and aspect ratio. In the shallow layer of the network, small objects are detected because they are divided into small grids as shown in Fig. 6(b), and in the deep layer of the network, large objects are detected because they are divided into large grids as shown in Fig. 6(c).

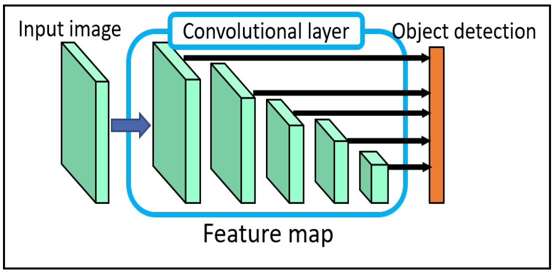
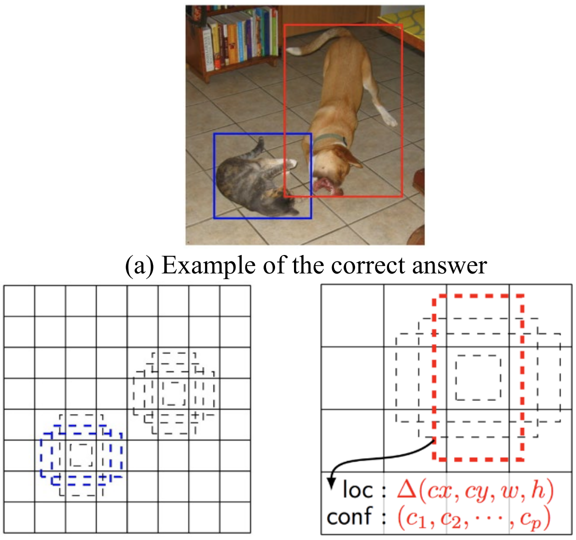


Figure 5: Structure of SSD


loc : $\Delta(cx, cy, w, h)$
conf : $(c_1, c_2, \cdots, c_p)$

(a) Example of the correct answer

(b) Feature map of shallow layers

(c) Feature map of deep layers

Figure 6: Searching with different feature maps

In this study, SSD512, which uses 512×512 images as input to the network is used. Compared to SSD300, the detection accuracy is improved, but the processing speed is reduced. Although there is a trade-off between accuracy and processing speed, in this study, SSD512 is chosen because this work emphasizes accuracy for accident prevention. In addition, a cropped part of the image from the captured image is used as the input image to increase the relative size of the human area and improve the accuracy. The coordinates of the intersection points are calculated from the lane information obtained in Section 2.1, and the image is cropped to a size of 512×512 and used as input to the SSD, Fig. 7.

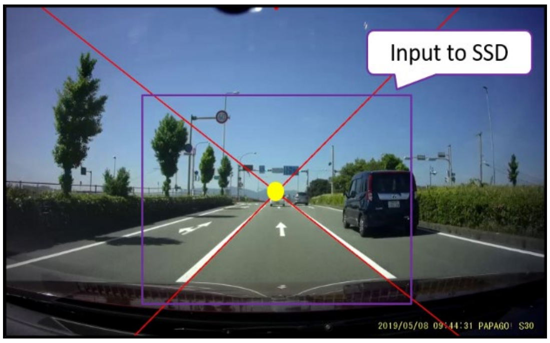


Figure 7: Image cropping

2.4.2 Tracking: In this paper, we propose a new method of object tracking using template matching to deal with the failure of SSD detection. In addition, when multiple persons are detected, it is difficult to match the persons in the current frame with those in the previous frame. However, to obtain information such as the direction of movement for each person, it is necessary to map each person to each frame. In our method, we use template matching for person tracking and mapping when multiple persons are detected.

If the number of persons detected in the current frame is less than the number of persons detected in the previous frame, we assume that the SSD failed to detect the persons and perform template matching to track the persons that could not be detected. If the number of people detected in the next frame is greater than or equal to the number of people detected in the previous frame, the system assumes that multiple people have been detected and compares the template image with the detected person's area for matching. When template matching is performed, the similarity is evaluated by SAD. Equation (2) is used to calculate the SAD. In eq. 2, *I(i,j)* is the image, and *T(i,j)* is the template.

$$SAD = \sum_i \sum_j |I(i,j) - T(i,j)| \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (2)$$

2.5 Person information acquisition

2.5.1 Location: To determine whether a person is in danger or not, it is important to know the location of the person. Therefore, we use the foot coordinates of a person to determine whether the person is on the sidewalk or in the lane.

First, we set the median of the bottom of the detected area of the person as the foot coordinates. Then, three points are calculated from the lane information obtained in Section 2.1. The intersection of the lanes, the intersection of each lane with the bottom edge of the image, and the area surrounded by the three points are defined as the lane area. Then, using the property of the outer product, the coordinates of the feet of the person are judged to be within the lane area.

2.5.2 Direction of movement: To determine whether a person is dangerous or not, it is important to recognize whether the person is moving left or right. However, it is difficult to analyze the movement of a person because the shooting position of an in-vehicle camera always changes. Therefore, we focus on the fact that a stationary object captured by an in-vehicle camera moves outward based on the vanishing point and estimate the direction of movement.

First, the intersection point is calculated from the lane obtained in Section 2.1, and its coordinates are set as the vanishing point. After that, we use the vanishing point to predict the direction of movement assuming that the person is stationary. After that, we obtain the direction in which the person moves and compare it with the predicted value to estimate which way the person is moving, regardless of the influence of the vehicle's movement, Fig. 8.

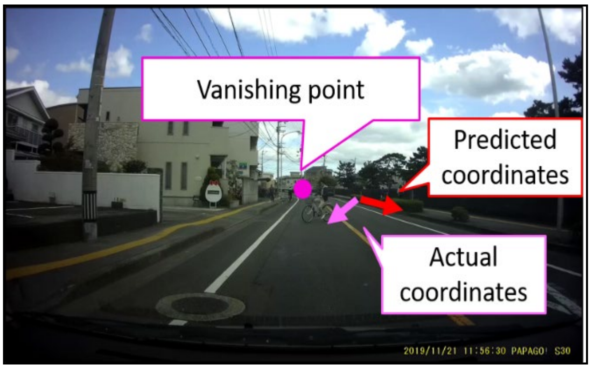
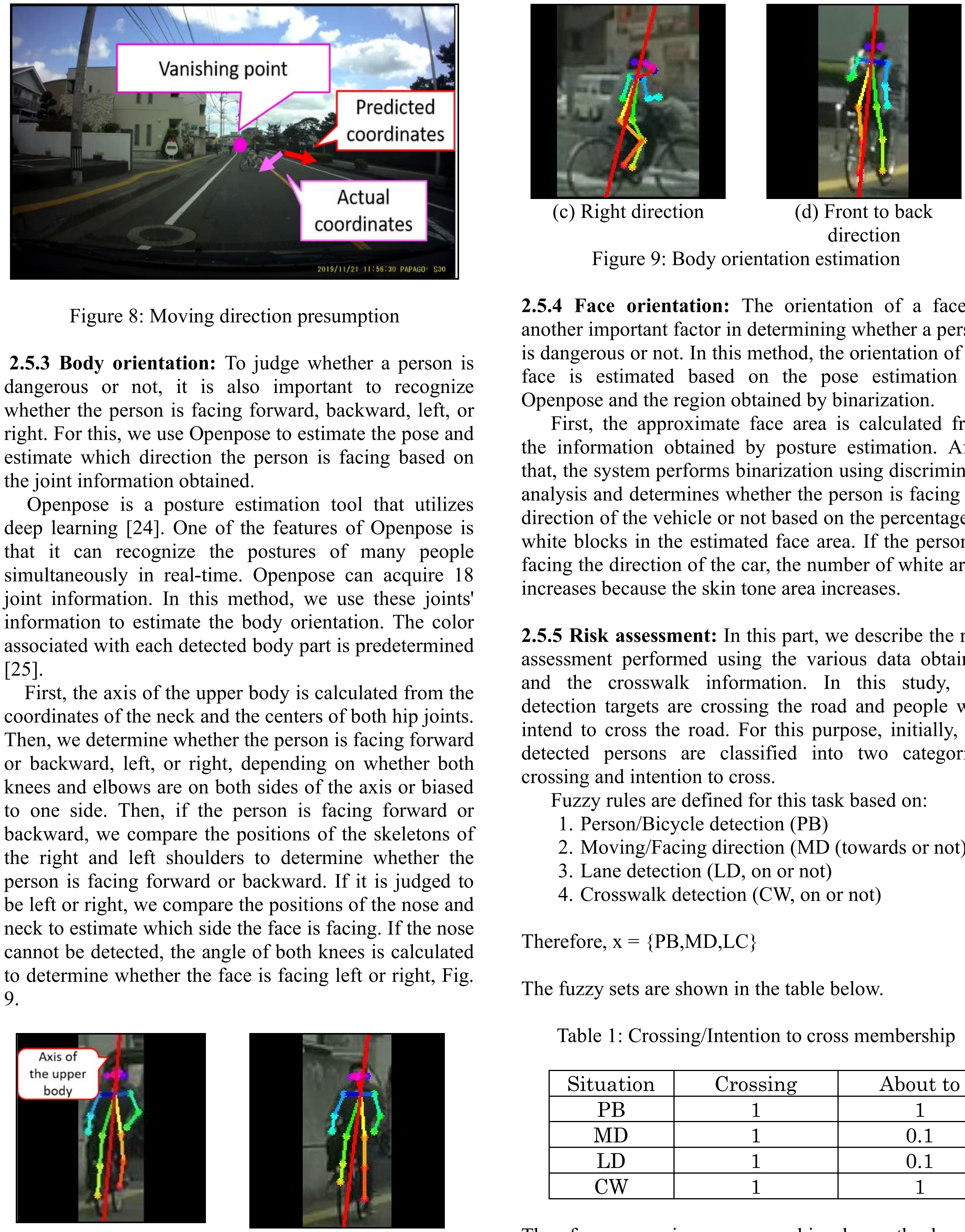


Figure 8: Moving direction presumption

2.5.3 Body orientation: To judge whether a person is dangerous or not, it is also important to recognize whether the person is facing forward, backward, left, or right. For this, we use Openpose to estimate the pose and estimate which direction the person is facing based on the joint information obtained.

Openpose is a posture estimation tool that utilizes deep learning [24]. One of the features of Openpose is that it can recognize the postures of many people simultaneously in real-time. Openpose can acquire 18 joint information. In this method, we use these joints' information to estimate the body orientation. The color associated with each detected body part is predetermined [25].

First, the axis of the upper body is calculated from the coordinates of the neck and the centers of both hip joints. Then, we determine whether the person is facing forward or backward, left, or right, depending on whether both knees and elbows are on both sides of the axis or biased to one side. Then, if the person is facing forward or backward, we compare the positions of the skeletons of the right and left shoulders to determine whether the person is facing forward or backward. If it is judged to be left or right, we compare the positions of the nose and neck to estimate which side the face is facing. If the nose cannot be detected, the angle of both knees is calculated to determine whether the face is facing left or right, Fig. 9.



(a) Left direction (b) Back direction

(c) Right direction (d) Front to back direction

Figure 9: Body orientation estimation

2.5.4 Face orientation: The orientation of a face is another important factor in determining whether a person is dangerous or not. In this method, the orientation of the face is estimated based on the pose estimation by Openpose and the region obtained by binarization.

First, the approximate face area is calculated from the information obtained by posture estimation. After that, the system performs binarization using discriminant analysis and determines whether the person is facing the direction of the vehicle or not based on the percentage of white blocks in the estimated face area. If the person is facing the direction of the car, the number of white areas increases because the skin tone area increases.

2.5.5 Risk assessment: In this part, we describe the risk assessment performed using the various data obtained and the crosswalk information. In this study, the detection targets are crossing the road and people who intend to cross the road. For this purpose, initially, the detected persons are classified into two categories: crossing and intention to cross.

Fuzzy rules are defined for this task based on:

1. Person/Bicycle detection (PB)
2. Moving/Facing direction (MD (towards or not))
3. Lane detection (LD, on or not)
4. Crosswalk detection (CW, on or not)

Therefore, x = {PB,MD,LC}

The fuzzy sets are shown in the table below.

Table 1: Crossing/Intention to cross membership

Situation	Crossing	About to
PB	1	1
MD	1	0.1
LD	1	0.1
CW	1	1

Therefore, a moving person or bicycle on the lane or crosswalk is considered crossing the road. Otherwise, if the person or bicycle is detected, are stationary, not on the lane or crosswalk, they are considered about to cross the road.

To estimate the risk posed, a new set of fuzzy rules are applied based on the subject intention categorized above and the car situation, especially the safe braking distance.

Fuzzy rules are defined for this task based on:

1. Crossing (CR)
2. Intend to Cross (IC)
3. Braking Distance (BR, enough or not)

Therefore, x1 = {CR,IC,BR}

The fuzzy sets are shown in the table below.

Table 2: Dangerous or not membership

Situation	Danger	No Danger
CR	1	1
IC	0.5	1
BR	0	1

If a subject is crossing or about to cross and the braking distance is not enough, then that is a dangerous situation, and the driver should be warned.

3. Experiment and Results

3.1 Environment

To evaluate the effectiveness of the proposed method, we conducted experiments on 89 scenes, including those with and without danger, which were captured while driving on the road. The resolution of the images acquired from the camera is 1280 × 720 and the frame rate is 30 fps. An example of the captured images is shown in Figure 10.

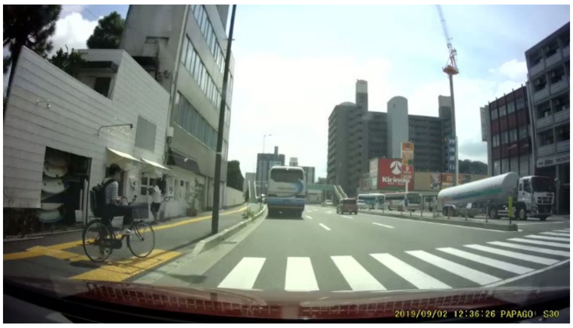
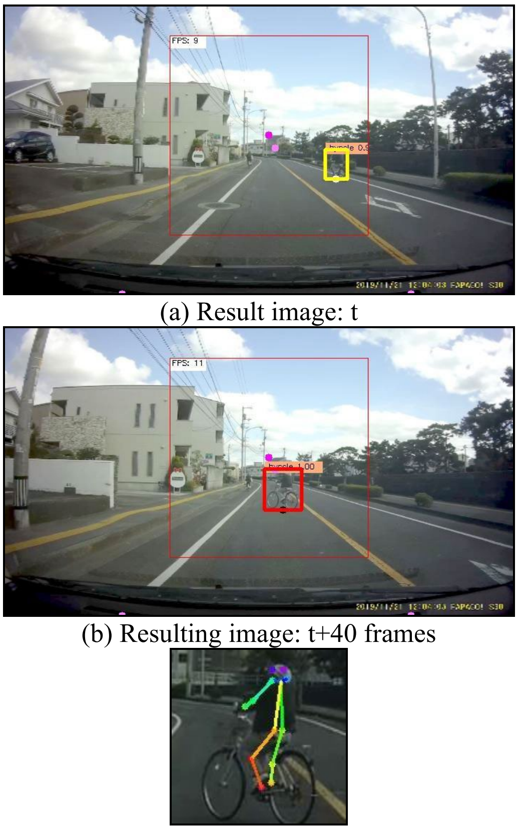


Figure 10: Experimental environment

3.2 Results

We show the processing results and the overall accuracy of two scenes out of the results of processing the captured images. In this experiment, a red rectangle is drawn when the person is judged to be crossing, and a yellow rectangle is drawn when the person is judged to be intending to cross.

In scene 1, a cyclist is crossing in front of the car on the right side of the road, moving in the same direction as the car. Figure 10 shows the results of the processing for each frame, and Table 3 shows the information obtained for the frame. In this scene, the timing at the beginning of the crossing is correctly processed as "crossing intention" and the timing during the crossing is correctly processed as "crossing.



(a) Result image: t

(b) Resulting image: t+40 frames

(c) Pose estimation result

Figure 11: Experimental results for scene 1

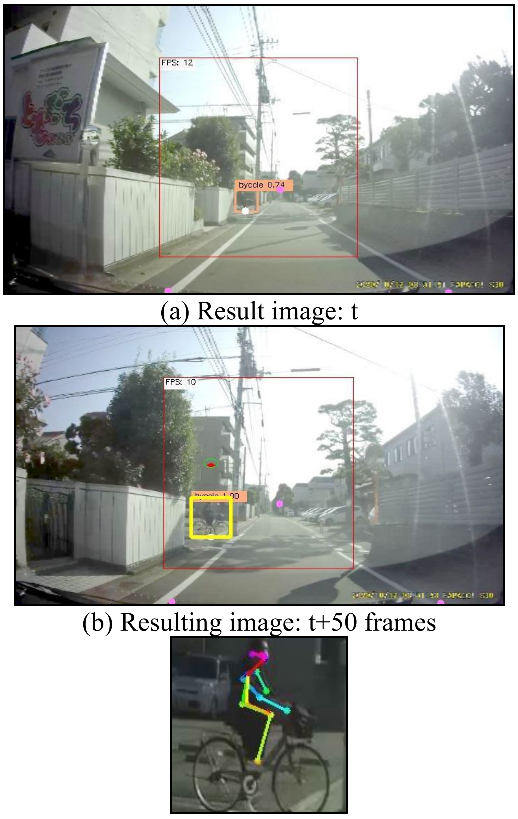
Based on our camera, the furthest the pedestrian (in the front view of the camera) can be is about 20m to ensure an OpenPose detection accuracy of above 93%. This means that our system is tested on vehicles driving at less than 30km/hr. According to [26], the stopping distance for a vehicle traveling at 30kph is 23 meters. SSD-VGG uses image size 512x512 because we selected the SSD512 model. After the detection of pedestrians or bicycles, their size is used as the input to the OpenPose algorithm. It should be noted that the maximum input size is, therefore, 512x512 assuming that to be the size of the detected person or bicycle. The smallest size to ensure accurate pose estimation is about 100x100 pixels.

In scene 7, the cyclist on the left is waiting to cross the street. Figure 12 shows the processing results for each frame, and Table 4 shows the acquired information for a particular frame.

Table 3: Information obtained at t+40 frames (LOC: Subject location, DM: Movement Direction, BD: Body direction, CW: Crosswalk, FD: Face Direction)

LOC	DM	BD	CW	FD
ON Lane	left	left	none	Back

Figure 12 shows the results for each frame, and Table 3 shows the acquired information for a particular frame. In this scene, the body orientation and face direction are correctly acquired, and thus the intention to cross can be correctly judged.



(a) Result image: t

(b) Resulting image: t+50 frames

(c) Pose estimation result

Figure 12: Experimental results for scene 7

Table 4: Information obtained at t+50 frames

LOC	DM	BD	CW	FD
OFF Lane	Right	Right	none	Front

The experiments were conducted on 7 scenes. The summarized results of pedestrian and bicycle detections are shown in Table 4. The accuracy is 87% and 91% respectively for pedestrian bicycles. The reason for the difference in accuracy could be due to the distance to the camera and object size. The bicycles were more likely to be near the vehicle and of bigger size at any distance compared to the pedestrians. However, once detected, there was no notable difference in the OpenPose accuracy for the pedestrians or bicycles.

Table 5: Search accuracy

	Number of people	Detected	Accuracy
Pedestrian	54	47	87
Bicycle	117	110	94
All	171	157	91

Table 5 shows the summarized results of the processing for all scenes for hazards based on the fuzzy rules generated.

Table 5: Accuracy of hazard determination

	Safe	Danger
Safe	84	9
Danger	4	64
Recall	0.94	
Precision	0.87	

Table 5 shows that the detection accuracy of bicycles is higher than that of pedestrians. This may be because the image size of the bicycle is larger, and the features can be extracted more easily by convolution. As shown in Table 5, both the reproduction rate and the fit rate were good in terms of judgment accuracy. This may be because the combination of multiple pieces of information can compensate for the failure in obtaining a specific piece of information by using other pieces of information.

4.0 Conclusion

In this paper, we proposed a method to detect and track a person in an image captured by an in-vehicle camera, and to detect a person who is crossing or intends to cross by using the acquired information such as position, and body orientation, and direction of movement.

Future work includes further improvement of the accuracy of person detection and tracking, improvement of danger judgment, further acquisition of a person's information, improvement of processing speed, and support for curved road sections.

References

1. The Japan Times, "Road Deaths in Japan fell to record low in 2020 amid COVID-19 stay home requests." https://www.japantimes.co.jp/news/2021/06/15/national/japan-traffic-accidents-covid-19/ (Retrieved 2022-1-20)

2. C. Papageorgiou and T. Poggio, "A Trainable Pedestrian Detection system", *International Journal of Computer Vision* (IJCV), pages 1:15–33, 2000
3. N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR), pages 1:886–893, 2005
4. A. Solichin and A. Harjoko, "A survey of Pedestrian Detection in video", *International Journal of Advanced Science and Application* (IJACSAI), pages 41–47, 2014
5. Bo Wu and Ram Nevatia, "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors", *IEEE International Conference on Computer Vision* (ICCV), pages 1:90–97, 2005
6. Mikolajczyk, K. and Schmid, C. and Zisserman, A. "Human detection based on a probabilistic assembly of robust part detectors", *The European Conference on Computer Vision* (ECCV), volume 3021/2004, pages 69–82, 2005
7. S. Piérard, A. Lejeune, and M. Van Droogenbroeck. "A probabilistic pixel-based approach to detect humans in video streams" *IEEE International Conference on Acoustics, Speech and Signal Processing*(ICASSP), pages 921–924, 2011
8. F. Fleuret, J. Berclaz, R. Lengagne and P. Fua, Multi-Camera People Tracking with a Probabilistic Occupancy Map*, IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, Nr. 2, pp. 267–282, February 2008.
9. R. Girshick, J. Donahue, T. Darrell et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," *in Proceedings Of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, 2014.
10. K. He, G. Gkioxari, P. Dollar, et al., "Mask r-cnn," *in Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, Long Beach, CA, USA, 2017.
11. R. Girshick, "Fast r-cnn," *in Proceedings Of the IEEE International Conference on Computer Vision*, pp. 1440–1448, Barcelona, Spain, 2015.
12. S. Ren, K. He, R. Girshick et al., "Faster R-CNN: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 39, no. 6, pp. 91–99, 2015.
13. J. Redmon, S. Divvala, R. Girshick et al., "You only look once: unified, real-time object detection," in *Proceedings Of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, Las Vegas, NV, USA, 2016.
14. J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR*), Honolulu, HI, USA, 2017.
15. J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, http://arxiv.org/abs/1804.02767
16. W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector*," in Proceedings of the European Conference on Computer Vision*, pp. 21–37, Springer, 2016.
17. W. Liu, S. Liao, W. Hu et al., "Learning efficient single-stage pedestrian detectors by asymptotic localization fitting," *in Proceedings Of the European Conference on Computer Vision*, pp. 643–659, Munich, Germany, 2018.
18. Z. Zheng, P. Wang, W. Liu et al., Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression, 2019, https://arxiv.org/abs/1911.08287.
19. B. A. Wang and C. W. Liao, Optimal Speed and Accuracy of Object Detection, 2020, https://arxiv.org/abs/2004.10934.
20. Di Tian,Yi Han, Biyao Wang, Tian Guan, and Wei Wei, A Review of Intelligent Driving Pedestrian Detection Based on Deep Learning, *Computational Intelligence and Neuroscience*, 1-16, 2021
21. ASV (Advanced Safety Vehicle), chrome-extension://ieepebpjnkhaiioojkepfniodjmjjihl/data/pdf.js/web/viewer.html?file=https%3A%2F%2Fwww.mlit.go.jp%2Fjidosha%2Fanzen%2F01asv%2Fdata%2Fasv5pamphlet-e.pdf (referenced 2020-12-19)
22. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg:" SSD: Single Shot MultiBox Detector", European conference on computer vision, Springer International Publishing, 978-3-319-46448-0, 10.1007/978-3-319-46448-0_2, pp.21–37 (2016).
23. Simonyan, K. and Zisserman, A: "Very deep convolutional Networks for Large-Scale Image Recognition", CVPR, arXiv:1409.1556, (2014)
24. Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh:OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 172-186, vol. 43, Jan. 2021.
25. OpenPose Experimental models, https://github.com/CMU- Perceptual-Computing-b/openpose_train/blob/master/experimental_models/README.md (referenced 2020-12-19)
26. Braking Distance, Wiki, https://en.wikipedia.org/wiki/Braking_distance. (referenced 2020-12-12)
27. H. Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao. Deep Ordinal Regression Network for Monocular Depth Estimation. In IEEE Computer Vision and Pattern Recognition (CVPR), 2018.
28. C. Godard, O. Mac Aodha, and G. J. Brostow. Unsupervised Monocular Depth Estimation with Left-Right Consistency. In International Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
29. A.Zaheer, M. Rashid, M. A. Riaz, and S. Khan. Single-View Reconstruction using orthogonal line-

pairs. Computer Vision and Image Understanding, 172:107–123, 2018.

Conflict of Interest

The authors declare that they have no conflict of interest.

Karungaru, Stephen, graduated from the Department of Electrical/ Electronics, Moi university in 1993. In 2001 he graduated from the Department of Information Science and Intelligent systems, Faculty of Engineering, Tokushima University with a master's degree. In 2004, he completed his Ph.D. work at the Department of Information Science and Intelligent systems, Faculty of Engineering, Tokushima University.

Ryosuke Tsuji graduated from the Department of Information Science and Intelligent systems, Faculty of Engineering, Tokushima University with a master's degree in 2021.

Terada, Kenji graduated from the Faculty of Science and Technology, Keio University in 1990. In 1992 he attained a master's degree from the Faculty of Science and Technology, Keio University. He obtained his Ph.D. degree from the Faculty of Science and Technology, Keio University in 1995.