KICS The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 9 (2023) 222–227

www.elsevier.com/locate/icte

A sensor fusion system with thermal infrared camera and LiDAR for autonomous vehicles and deep learning based object detection

Ji Dong Choi, Min Young Kim*

Department of Future Automotive and IT Convergence, Kyungpook National University, Daegu, Republic of Korea
School of Electronic and Electrical Engineering, Kyungpook National University, Daegu, Republic of Korea

Received 12 September 2021; received in revised form 30 October 2021; accepted 29 December 2021
Available online 5 January 2022

Abstract

Vision, Radar, and LiDAR sensors are widely used for autonomous vehicle perception technology. Especially object detection and classification are primarily dependent on vision sensors. However, under poor lighting conditions, dazzling sunlight, or bad weather an object might be difficult to be identified with general vision sensors. In this paper, we propose a sensor fusion system that combines a thermal infrared camera and a LiDAR sensor that can reliably detect and identify objects even in environments with poor visibility, such as day or night. The proposed method obtains the external parameters of the two sensors by designing and manufacturing a 3D calibration target to externally calibrate the thermal infrared camera and the LiDAR sensor. To verify the performance, experiments were conducted in day and night environments. The proposed sensor system and fusion algorithm show that it can reliably detect and identify objects even in environments with poor visibility, such as day or night.

Keywords: Sensor fusion; LiDAR; Thermal infrared camera; Autonomous vehicles; Object detection; Convolution neural network

1. Introduction

Object recognition technology for autonomous vehicles has been extensively studied for many years [1–3]. Recognizing the environment around an autonomous vehicle is the first step in autonomous driving and a key factor in determining its performance. This is because the purpose of recognition is to detect obstacles in various driving environments and provide a driving route. Object recognition technology is divided into communication type and sensor type. The communication type recognizes the surrounding environment through communication with the outside using communication equipment mounted on the vehicle. The communication method between the transmission equipment inside the vehicle and the receiving equipment of the vehicle and a specific object is called Vehicle to Everything (V2X). Although V2X communication has a longer range compared to the sensor type, it is vulnerable to hacking, and its performance varies depending on the surrounding environment or the state of the autonomous vehicle.

The low antenna height and high mobility of V2X communication greatly change the communication channel and propagation path environment. A lot of research has been done over the years to improve the performance of V2X communication. Among them, there are communication standard technologies such as Wireless Access in Vehicular Environment (WAVE), Long Term Evolution (LTE), and 5G. Research on WAVE communication is stagnant due to the limitations of IEEE 802.11a/g wireless LAN technology specifications. In addition, since previous studies focused on performance evaluation through simulation [4], they did not accurately reflect the environment around the actual autonomous vehicle. If accurate communication channel performance estimation is not possible, the reliability of the autonomous vehicle communication system design cannot be guaranteed. As a result, problems arise in autonomous vehicle communication services and have serious consequences, threatening the driving safety of autonomous vehicles.

The sensor type uses a vision sensor mounted on an autonomous vehicle to detect surrounding objects or obstacles [5]. However, each sensor has its pros and cons, and it only works in a limited environment or within the sensing range of the sensor. Existing autonomous vehicle recognition research uses Vision, Radio Detection and Ranging (Radar),

* Corresponding author at: School of Electronic and Electrical Engineering, Kyungpook National University, Daegu, Republic of Korea.
E-mail addresses: jidong1119@naver.com (J.D. Choi), minykim@knu.ac.kr (M.Y. Kim).
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2021.12.016

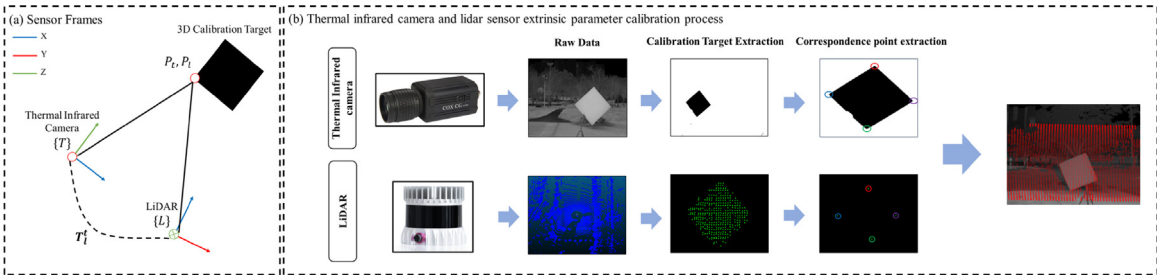

Fig. 1. Sensor frame and thermal infrared camera and extrinsic parameter calibration method of LiDAR sensor. (a) The x-axis, y-axis, z-axis of the sensor frames are shown in red, green and blue color, (b) Thermal infrared camera and LiDAR sensor extrinsic parameter calibration process. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and Light Detection And Ranging (LiDAR) sensors to detect objects.

Radar examines radio signals and analyzes radio energy reflected from obstacles. Radar sensors are cheaper than LiDAR sensors and mainly determine the presence of obstacles [6]. LiDAR sensors can acquire approximately 360° three-dimensional spatial information around an autonomous vehicle by irradiating light signals in a specific way and analyzing the light energy reflected from obstacles. These sensors are used for 3D mapping, localization, and object detection.

On the other hand, thermal infrared cameras reliably capture objects in low-visibility and high-contrast conditions such as at night, shadows, sunsets, and sunrises, in situations with severe glare from direct sunlight or car headlights, and also in environments with poor visibility such as fog or smoke.

Information from a single sensor alone cannot guarantee the reliability of cognitive technologies in complex autonomous vehicle environments. That is why autonomous vehicles are equipped with multiple sensors to improve their perception of their surroundings. In order to increase the reliability of recognition, the data acquired from the camera and LiDAR sensor must be matched to the spatiotemporal coordinate frame. Time matching between different sensors can be said to be exact synchronization between all the sensors they use, and common coordinate frame matching can be said to be a geometric correction of extrinsic parameters between the three-dimensional coordinate systems of the sensors.

Studies on extrinsic parameter calibration methods for cameras and LiDAR sensors require specific 3D calibration targets [7,8]. [7] proposed a method to calibrate extrinsic parameters between the camera and LiDAR sensor using a 3D calibration target with a circular hole in the plane. The coordinate system between the two sensors was matched through the center coordinates extracted by detecting the circles from the two sensors. In [8], the two sensors were matched through the coordinates of the extracted marker by attaching a special marker to the plane. Although research on external parameter correction between a vision camera and a LiDAR sensor has been actively conducted, there is still no direct external parameter correction method between a thermal infrared camera and a LiDAR sensor.

It is not without research on calibration of thermal infrared cameras and LiDAR sensors. In [9], a study on external parameter correction using a thermal infrared camera, a visual camera, and a LiDAR sensor was conducted. However, as actual image cameras are essential, unnecessary computations have increased. In this paper, we propose an object detection algorithm robust to day and night environments for autonomous vehicles developed by the direct external infrared parameter correction algorithm [10] between thermal infrared cameras and LiDAR sensors.

The paper is organized as follows: Section 2 examines the process of extracting key points using the characteristics of thermal infrared cameras and LiDAR sensors and calibration of extrinsic parameters. Section 3 analyzes the performance of the proposed method using the data acquired in the experimental environment in day and night environments. Section 4 describes the conclusions and concludes this paper.

2. Calibration of extrinsic parameters of thermal infrared cameras and LiDAR sensors

Fig. 1(a) shows the sensor frame. The transformation matrix from the LiDAR frame $\{L\}$ to the thermal infrared camera frame $\{T\}$ is T_l^t. T_l^t can be expressed as follows.

$$T_l^t = \begin{bmatrix} r_{11} & r_{12} & r_{31} & t_1 \\ r_{21} & r_{22} & r_{32} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \tag{1}$$

where r is the rotation matrix and t is the translation matrix.

2.1. Thermal infrared camera calibration

Convergence studies of cameras and LiDAR sensors are being actively conducted. However, compared to visual cameras, thermal infrared cameras cannot extract the exact shape of the 3D calibration target, so direct extrinsic parameter calibration is difficult. To calibrate the extrinsic parameters of the thermal infrared camera and the LiDAR sensor, first, the intrinsic parameters of the thermal imaging camera must be obtained. Fig. 1(b) shows a processor that extracts the feature points from a thermal infrared camera and a LiDAR sensor and performs extrinsic parameter calibration.

In this paper, Zhang's calibration algorithm was used [11]. Zhang's calibration algorithm uses a checkered pattern on a flat 3D calibration target. Convenient, easy to use and provides accurate calibration results. However, with a thermal infrared camera, using a circular pattern rather than a checkered pattern

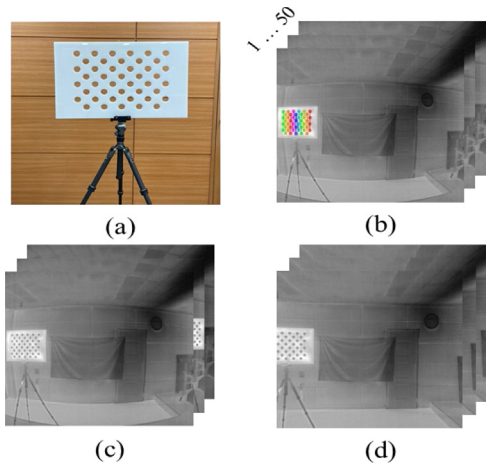

Fig. 2. Thermal infrared camera calibration process. (a) Thermal camera 3D calibration target, (b) Distorted thermal image, (c) Feature point extraction result for distortion calibration, (d) Undistorted thermal image.

may yield more accurate results. In order to utilize the characteristics of the thermal infrared camera, a three-dimensional calibration target made of acrylic was designed and manufactured as shown in Fig. 2(a). As shown in Fig. 2(b), a more accurate prototype can be extracted from the thermal image by heating the 3D calibration object using a heating gun. The extracted circular feature points are shown in Fig. 2(c). The result of removing the distortion of the thermal infrared image is shown in Fig. 2(d).

2.2. Thermal image feature point extraction for extrinsic parameter calibration

In order to utilize the information of two different sensors as one fusion system, extrinsic parameter calibration that combines the coordinate system between the sensors into one common coordinate frame is required. There is no direct extrinsic parameter calibration method for thermal infrared cameras and LiDAR sensors. Zhang's method obtains the conversion matrix between the visual camera and the LiDAR sensor, and then the conversion matrix between the thermal infrared camera and the visual camera. Finally, the extrinsic parameters between the LiDAR and the thermal infrared camera were calculated by multiplying the above two matrices [11].

In this paper, we propose an algorithm for the direct extrinsic parameter calibration of thermal infrared cameras and LiDAR sensors. First, the entire thermal image brightness distribution is analyzed using a histogram, which is a technique for analyzing the frequency of image brightness values in a thermal image. As shown in Fig. 3, the 3D calibration target is separated through the value corresponding to the upper n of the histogram brightness value. where n is the distribution ratio of the upper brightness value. The image processing technique, binarization, is applied to the thermal image $t(x, y)$ using the histogram upper n values to isolate the 3D calibration target:

$$t(x,y) = \begin{Bmatrix} 255 & if \quad t(x,y) > n \\ 0 & otherwise \end{Bmatrix} \tag{2}$$

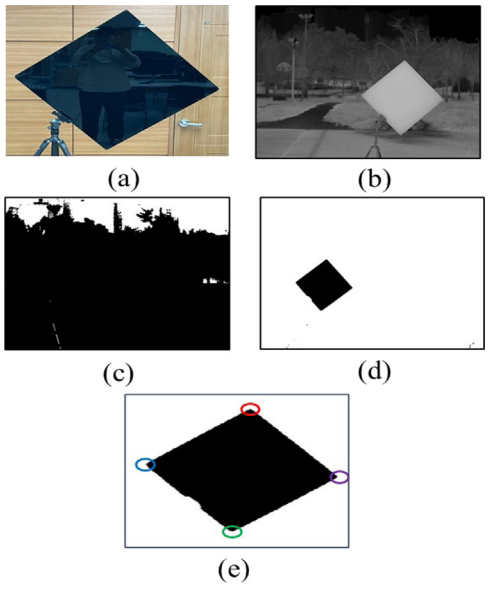

Fig. 3. Thermal image threshold processing using histogram. (a) Thermal infrared camera and LiDAR extrinsic 3D calibration target, (b) Original thermal image, (c) Top 10% result, (d) Top 70% result, (e) Correspondence point extraction.

From the extracted 3D calibration target, a corresponding point is extracted as shown in Fig. 3(e). By comparing the coordinate values of pixels in the thermal image $t\ (x, y)$ through $t_{max}\ (t_x, t_y)$, $t_{min}(t_x, t_y)$, a total of four corresponding points are extracted as:

$$t_{max}\ (x,y) = \begin{Bmatrix} max(t_x, t_y) & max\ (x,y) > t(x,y) \\ max\ (x,y) & otherwise \end{Bmatrix} \tag{3}$$

$$t_{min}\ (x,y) = \begin{Bmatrix} min(t_x, t_y) & min\ (x,y) < t(x,y) \\ min\ (x,y) & otherwise \end{Bmatrix} \tag{4}$$

2.3. LiDAR 3D point cloud feature point extraction for extrinsic parameter calibration

A 3D point cloud of the LiDAR sensor is shown in Fig. 4(a). However, in order to fit the field of view (FOV) of the thermal infrared camera without using the full 3D point cloud, primary filtering was performed leaving only the front 180° as shown in Fig. 4(b). The 3D point cloud data corresponding to the 3D correction target is extracted through secondary filtering according to the distance value. Then, the three-dimensional calibration target is extracted using the RANSAC algorithm [12]. The optimal plane is calculated using internal values determined through the RANSAC algorithm. By comparing the 3D point cloud coordinate values in the 3D point cloud $l(x, y, z)$ corresponding to the 3D correction target, a total of four corresponding points $l_{max}(l_x, l_y, l_z)$, $l_{min}(l_x, l_y, l_z)$ Extracting the four corresponding points is as follows.

$$l_{max}\ (x,y,z) = \begin{Bmatrix} max(l_x, l_y, l_z) & max\ (x,y,z) > l(x,y,z) \\ max\ (x,y,z) & otherwise \end{Bmatrix} \tag{5}$$

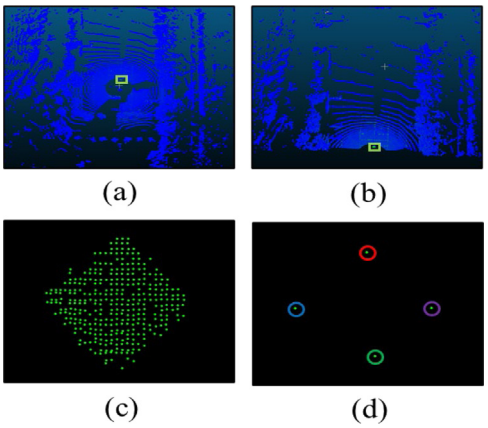

Fig. 4. LiDAR point cloud extraction result. (a) Raw point cloud, (b) Front 180° area extraction result, (c) Extract 3D calibration targets, (d) Results of feature point extraction from the 3D calibration targets.

Fig. 5. Autonomous vehicle demonstrating the proposed algorithm.

$$l_{min}(x, y, z) = \begin{cases} min(l_x, l_y, l_z) & min(x, y, z) < l(x, y, z) \\ min(x, y, z) & otherwise \end{cases} \tag{6}$$

where $l(x, y, z)$ means the 3D coordinate system axis of the LiDAR sensor.

2.4. Calibration of extrinsic parameters of thermal infrared cameras and LiDAR sensors

In Section 2.1, we obtained the intrinsic parameters of the thermal imaging camera. If the unique parameters of the thermal infrared camera are known, the $R|t$ between the world coordinate system of LiDAR sensor and the image coordinates of the thermal infrared camera is estimated using the PnP algorithm [13] implemented in the OpenCV library. Here, R is a matrix that converts the 3D world coordinate system of the LiDAR sensor to the 2D image coordinate system of the thermal infrared camera, R is the rotation matrix, and t is the translation matrix. The estimated relationship between the thermal infrared camera and the LiDAR sensor is as follows:

$$s\begin{bmatrix} T_u \\ T_v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{31} & t_1 \\ r_{21} & r_{22} & r_{32} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} L_x \\ L_y \\ L_z \\ 1 \end{bmatrix} \tag{7}$$

where (T_u, T_v) is the image pixel coordinates, (f_x, f_y) is a focal length, (c_x, c_y) is an optical center, r is a rotation matrix, t is a transformation matrix, and (L_x, L_y, L_z) is a coordinate system of the LiDAR sensor.

Fig. 6. Thermal image object detection results.

3. Experiment result

3.1. Configuration of autonomous vehicle system

As shown in Fig. 5, the sensor consists of a COX CG-640 thermal infrared camera (resolution: 640 × 480), a Logitech Brio visual camera (resolution: 640 × 480), and an OUSTER OS-1 LiDAR, mounted on the roof of an autonomous vehicle. All algorithms run on desktop, NVIDIA TITAN RTX D6 24 GB, intel® Core i9 9900k CPU @ 3.60 GHz, 32 GB RAM, OS used Ubuntu 18.04 and ROS melodic.

3.2. YOLOv4 for object detection in thermal images

Studies on learning and cognition based on a lot of data for object detection in autonomous vehicles are being actively conducted. In this paper, objects are detected by dividing them into obstacles, vehicles, and pedestrians that can be dangerous factors when an autonomous vehicle drives on a road. For object detection, YOLOv4, one of the Convolutional Neural Network (CNN) models, was applied. YOLOv4 is a representative one-stage-detector algorithm and has the advantage of fast detection because localization and classification are performed simultaneously. The result of object detection is shown in Fig. 6.

3.3. Object detection using thermal infrared camera and LiDAR sensor fusion

The purpose of the thermal infrared camera and LiDAR sensor fusion proposed in this paper is to increase the accuracy of object detection by fusion of sensors with different advantages. In Section 2, the coordinate system of the LiDAR sensor was matched with the thermal infrared camera through extrinsic parameter calibration. Objects can be detected through distance values by reprojecting the 3D point cloud data of LiDAR in the bounding box detected in the thermal infrared image, but the accuracy is lowered when there are many objects to be detected in the thermal image. Accuracy can be improved by detecting objects in two spaces: an image from a thermal infrared camera and a 3D point cloud from a LiDAR sensor.

(C_x, C_y) which is the center point of the bounding box of the object detected in the thermal image is obtained. The center

point (C_x, C_y) of the obtained 2D bounding box is converted back to the 3D world coordinate system using $R|t$ obtained through extrinsic parameter calibration of the two sensors, and then the object is detected using the 3D bounding box specified in advance. The detected results can be seen in Fig. 7.

3.4. Quantitative performance verification using real data

Real data were acquired with the autonomous vehicle as shown in Fig. 5. To verify the performance of the method proposed in this paper, data from vision cameras, thermal infrared cameras, and LiDAR sensors were acquired in both day and night environments. In this paper, verification data was acquired in the following environment; (1) City area and (2) Kyungpook National University campus.

To evaluate the quantitative performance of the thermal infrared camera and LiDAR sensor-based object detection algorithm proposed in this paper, a performance comparison experiment was conducted between the visual camera and the LiDAR extrinsic parameter calibration algorithm. In [14], a calibration method for a visual camera and a LiDAR sensor is proposed. The checkered pattern is extracted from the visual camera and the edge of the 3d calibration target is extracted from the LiDAR sensor through plane fitting to estimate the external parameters of the two sensors.

In object detection, research to improve accuracy is mainly conducted. Accuracy is primarily achieved through comparisons between the correct answer values and the results predicted by the model. The high accuracy of the model means that the model regresses a bounding box similar to the correct answer, classifying the types of objects within the box. In general, object detection studies are considered undetected if the type of object cannot be predicted.

Precision is used with recall. It is an indicator that can indicate how accurate the predicted result is. It is possible to know how accurate the detection result is because it is possible to know what percentage of the detected objects got the correct answer value.

Precision is the predicted ratio of positive detections for all positive detection samples, which is the ratio of true positives (TP) to total positive results $(TP + FP)$:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall is the ratio of the positive prediction in the total positive samples of the Ground truth, which is the ratio of Positive prediction (TP) in total ground data $(TP + FN)$:

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

In general, precision and recall are inversely related to each other. Increasing recall increases false positives and decreasing false positives decreases recall. Therefore, precision–recall curves are used to properly compare and evaluate performance. The PR curve (Precision–Recall Curve) is one of the methods to evaluate the performance of an object detection algorithm, and the precision and recall values also vary according to the change of the threshold. Typically, an object detection model assigns a threshold and only detects when a certain value is exceeded. Fig. 8 shows the PR curve result of the algorithm proposed in this paper. In the result graph, it can be divided into situation A and situation B. In situation A, precision and recall are high. In case B, the classification of objects was exactly the same, but all objects were not detected, so it can be seen that the precision and recall are lower than in case A. In this paper, the threshold is set to 60%.

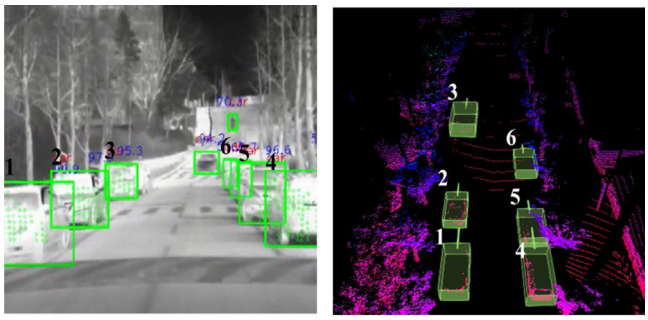

Fig. 7. Thermal infrared camera and LiDAR sensor fusion detection result. (a) Thermal image object detection result, (b) 3D point cloud object detection result.

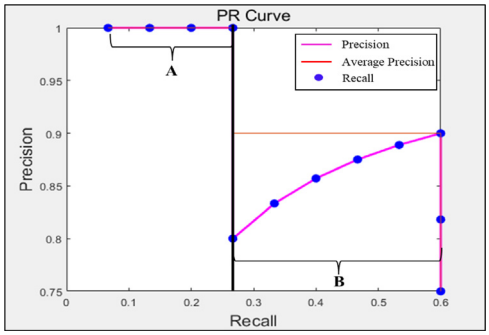

Fig. 8. Precision–Recall curve results graph.

The PR curve is easy to grasp the overall performance of an algorithm, but there is a limit to quantitatively verify the performance of two different algorithms. Therefore, recently, we compare the performance of recognition algorithms with Average Precision (AP). The higher the AP, the better the overall performance of the proposed algorithm. In the field of computer vision, the performance of object detection and image classification algorithms is evaluated by AP.

As shown in Fig. 9, the object detection AP of the daytime environment visual camera and LiDAR sensor was 56.167%, and the object detection AP of the thermal infrared camera and LiDAR sensor was 55.914%.

Table 1 shows the object detection results of the visual camera, LiDAR sensor, thermal infrared camera, and LiDAR sensor. It shows similar performance in the daytime environment, but the visual camera and LiDAR sensor object detection AP dropped to 49.878% in the night environment. This is because visible cameras cannot identify objects in nighttime environments with little or no lighting. On the other hand,

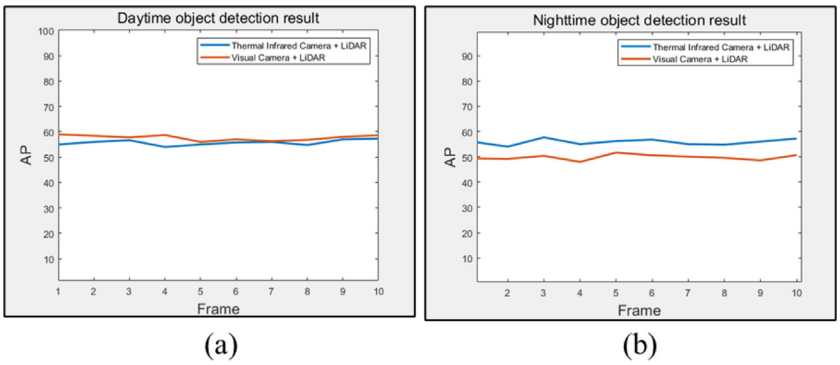

Fig. 9. Object detection AP verification result. (a) Daytime, (b) Nighttime.

Table 1
Comparison of object detection precision, recall and AP values of visual camera and LiDAR sensor algorithm and thermal infrared camera and LiDAR sensor algorithm.

Index	Visual camera + LiDAR/Thermal infrared camera + LiDAR		
	Precision (%)	Recall (%)	AP (%)
Daytime	89.341/ 86.872	86.711/ 84.161	56.167/ 55.914
Nighttime	74.672/ 92.471	71.431/ 89.412	49.878/ 57.516

thermal infrared camera and LiDAR showed higher performance than visual camera and LiDAR at 57.516% in night environment.

4. Conclusion

We proposed a direct extrinsic parameter calibration method to obtain extrinsic parameters between a thermal infrared camera and a LiDAR sensor. A 3D calibration target was manufactured to match the characteristics of the thermal infrared camera and LiDAR sensor, and extrinsic parameters were calibrated. To verify the performance, the accuracy and recall were estimated by acquiring real data of day and night environments. Through the estimated precision and recall, we can improve performance by setting a threshold to fit the model. In the future, 3D object detection research will be conducted using the reconstructed 3D thermal point cloud data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144).

References

[1] Y. Jamtsho, P. Riyamongkol, R. Waranusast, Real-time license plate detection for non-helmeted motorcyclist using YOLO, ICT Exp. 7 (1) (2021) 104–109, http://dx.doi.org/10.1016/j.icte.2020.07.008.

[2] J. Lu, H. Sibai, E. Fabry, D. Forsyth, NO need to worry about adversarial examples in object detection in autonomous vehicles, 2017, https://arxiv.org/abs/1707.03501.

[3] A. Bochkovskiy, C.Y. Wang, H.Y.M. Liao, YOLOv4 optimal speed and accuracy of object detection, 2020, https://arxiv.org/abs/2004.10934.

[4] J.H. Joo, M.C. Park, D.S. Han, V. Pejovic, Deep learning-based channel prediction in realistic vehicular communications, IEEE Access 7 (2019) 27846–27858.

[5] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, Vision meets robotics: The KITTI dataset, Int. J. Robot. Res. 32 (11) (2013) 1231–1237.

[6] G. Péter, B. Kiss, V. Tihanyi, Vision and odometry based autonomous vehicle lane changing, ICT Exp. 5 (4) (2019) 219–226, http://dx.doi.org/10.1016/j.icte.2019.09.005.

[7] F.S.A. Rodriguez, V. Fremont, Extrinsic calibration between a multi-layer lidar and a camera, in: International Conference on Multisensor Fusion and Integration for Intelligent Systems, MFI, Seoul, Korea, Aug 20-22 2008, 2008, pp. 214–219.

[8] P. An, T. Ma, K. Yu, B. Fang, J. Zhang, W. Fu, J. Ma, Geometric Calibration for LiDAR-Camera System Fusing 3D-2D and 3D-3D Point Correspondences, Vol. 28, OSA Publishing, 2020, pp. 2122–2141.

[9] J. Zhang, P. Siritanawan, Y. Yue, C. Yang, M. Wen, D. Wang, A two-step method for extrinsic calibration between a sparse 3d lidar and a thermal camera, in: 15th International Conference on Control, Automation, Robotics and Vision, ICARCV, 2018, Singapore, (2018) (1039) 18-21–1044.

[10] J.D. Choi, M.Y. Kim, A sensor fusion system with thermal infrared camera and LiDAR for autonomous vehicles: Its calibration and application, in: The 12th International Conference on Ubiquitous and Future Networks, 2021, pp. 361–365.

[11] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 1330–1334.

[12] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.

[13] V. Lepetit, F. Moreno-Nogure, P. Fua, EPnP: An accurate O(n) solution to the PnP problem, Int. J. Comput. Vis. 81 (2009) 155–166.

[14] G.H. Lee, J.D. Choi, J.H. Lee, M.Y. Kim, Object detection using vision and LiDAR sensor fusion for multi-channel V2X system, in: International Conference on Artificial Intelligence in Information and Communication, 2020.