A Sensor Fusion System with Thermal Infrared Camera and LiDAR for Autonomous Vehicles: Its Calibration and Application

Ji Dong Choi
Department of Future Automotive and IT Convergence
Kyungpook National University
Daegu, Korea
jidong1119@naver.com

Min Young Kim
School of Electronics Engineering
Kyungpook National University
Daegu, Korea
minykim@knu.ac.kr

***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 weathers an object might be difficult to be identified with general vision sensors. In this paper, we propose a sensor fusion system with a thermal infrared camera and LiDAR sensor that can reliably detect and identify objects even in environments where visibility is poor, such as in severe glare and fog or smoke. The proposed method obtains intrinsic parameters by calibrating the thermal infrared camera and LiDAR sensor. Extrinsic calibration algorithm between two sensors is made to obtain the extrinsic parameters (rotation and translation matrix) using 3D calibration targets. This system and proposed algorithm show that it can reliably detect and identify objects even in hard visibility environments, such as in severe glare due to direct sunlight or headlights or in low visibility environments, such as in severe fog or smoke.**

Keywords—Calibration; Autonomous Vehicles; Sensor fusion; LiDAR; Thermal Infrared Camera;

I. INTRODUCTION

Autonomous driving technology can be divided into perception, judgment, and control. Among them, perceiving the surrounding environment of an autonomous vehicle is the first step and a key determinant of performance. This is because the goal of perception is to detect the obstacles and provide the driving path in various driving environments. Perception technology is divided into communication type and sensor type. The communication type recognizes the surrounding situation through communication with the outside using communication equipment installed in 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 the V2X communication has a longer range than the sensor type, it is vulnerable to hacking and its performance varies depending on the surrounding environment or vehicle condition.

The sensor type uses sensors mounted on autonomous vehicles to detect nearby objects or obstacles. However, each sensor has its advantages and disadvantages, and it only works in a limited environment or within the sensing range of the sensor. The perception sensors for traditional autonomous vehicles are Vision, Radar, and LiDAR.

Fig. 1. Object detection in dazzling sunlight (top) and object detection in dark lighting environments (bottom).

LiDAR is a sensor that investigates optical signals in a certain way and analyzes light energy reflected from obstacles and acquires three-dimensional spatial information around autonomous vehicles. Thermal infrared cameras can perform even in low visibility and high contrast environments such as night, shadow, sunset, and sunrise, and low visibility such as direct sunlight or headlights, fog or smoke. Currently, autonomous vehicle perception research uses a vision sensor to detect objects. As shown in Fig. 1. in poor lighting conditions, dazzling sunlight, bad weather an object cannot be identified with a general Visual Camera. However, the thermal infrared camera can reliably detect objects even in this situation.

Information from a single sensor alone cannot guarantee the reliability of perception in complex autonomous vehicle surroundings. Therefore, autonomous vehicles are equipped with several sensors to improve perception rates in their surroundings. To combine information from different sensors into a common coordinate frame, we need to know the relative positions and directions between sensors. These parameters can be obtained by extrinsic calibration. This paper focuses on direct extrinsic calibration between LiDAR and thermal infrared camera.

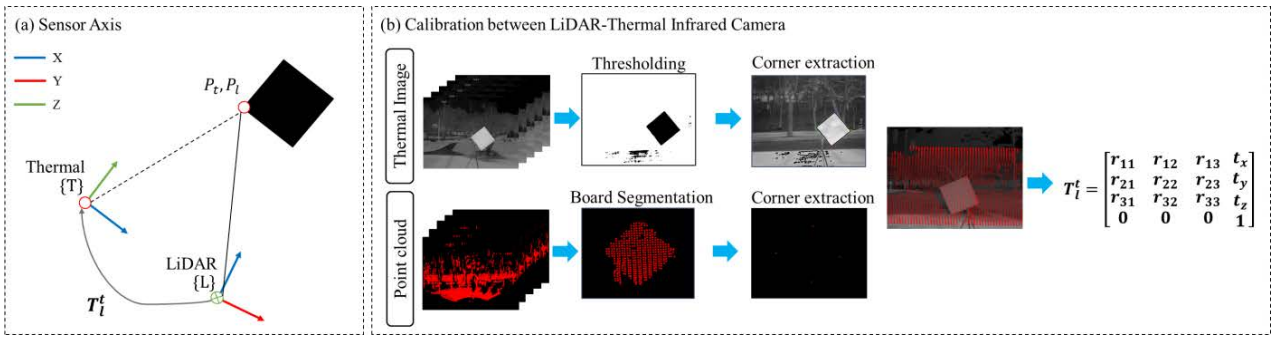

Fig. 2. Sensor frame and thermal infrared camera and extrinsic calibration method of LiDAR sensor. (a) The x-axis, y-axis, z-axis of the sensor frames are shown in red, green and blue color. (b) LiDAR point $P_l = \{X_l Y_l Z_l\}^T$ in the LiDAR frame $\{L\}$ can be transformed to $P_t = (X_t, Y_t, Z_t)^T$ in the thermal infrared camera frame $\{T\}$ by the transformation matrix T_l^t.

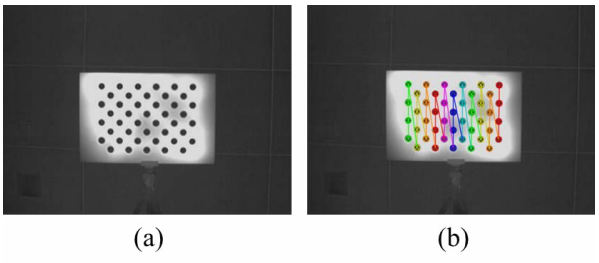

Fig. 3. Thermal infrared camera internal calibration results. (a) Thermal image original. (b) Feature point extraction results for intrinsic calibration.

II. Method

The relative position and orientation of the multiple sensors mounted on autonomous vehicles are unknown. This can be represented by the transformation matrix T. For example, the transformation matrix T_l^t from the LiDAR frame $\{L\}$ to the thermal infrared camera frame $\{T\}$ can be expressed as shown in (1).

$$T_l^t = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

As shown in Fig. 2a, LiDAR point $P_l = (x_l, y_l, z_l)^T$ in $\{L\}$ is given, then the point in the thermal infrared camera frame $\{T\}$, $P_t = (X_t, Y_t, Z_t)^T$, can be obtained by (2).

$$P_t = R_l^t P_l + T_{l;}^t \quad (2)$$

A. Thermal infrared camera calibration

To calibration the LiDAR and thermal infrared camera, first, the intrinsic parameters of the thermal infrared camera must be obtained. In this paper, we used the Zhang' calibration method implemented in OpenCV [1]. Zhang calibration algorithm uses a check pattern. It is convenient, easy to use, and provides accurate calibration results. The circle pattern is more accurate than the check pattern because the thermal infrared camera cannot extract the exact shape of the 3D calibration targets compared to the visual camera. In order to utilize the characteristics of the thermal infrared camera, the 3D calibration targets was heated as shown in Fig. 3a. Heated target can be more accurately extracted from thermal images. The extracted circular feature points are shown in Fig. 3b.

Since the 3D calibration targets is a flat board, $Z = 0$. Therefore, the projected image pixels and 3D calibration targets are as shown in (3). Finally, it can be summarized by (4). K is the camera intrinsic matrix, and $R|t$ is the rotation and tranlsation matrix from the world frame to the pixel frame.

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3)$$

$$= K[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (4)$$

Cheap small cameras generate significant image distortion. Major distortions include radiation distortion and tangential distortion. Radial distortion causes the image to bend. This phenomenon gets worse as we move away from the center of the image. Radial distortion can be solved using (5).

Tangential distortion occurs because the camera lens is not perfectly parallel to the image plane. Therefore, some areas of the image may appear closer than expected. Tangential distortion can be solved by using (6).

$$\begin{aligned} x_{distorted} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{distorted} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{aligned} \quad (5)$$

$$\begin{aligned} x_{distorted} &= x + [2p_1 xy + p_2(r^2 + 2x^2) \\ y_{distorted} &= y + [p_1(r^2 + 2y^2) + 2p_2 xy] \end{aligned} \quad (6)$$

In this paper, correction was performed with 50 test images, and finally, the distortion coefficient (7), the focal length (f_x, f_y) and optical center (c_x, c_y) were obtained (8).

$$Distortion\ coefficients = (k_1, k_2, p_1, p_2, k_3,) \tag{7}$$

$$camera\ matrix = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

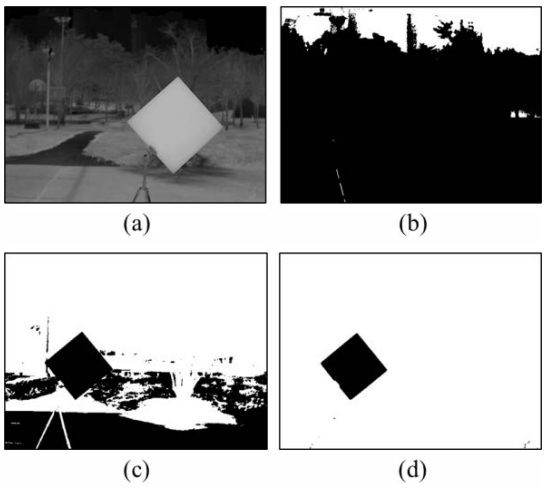

Fig. 4. Thermal image threshold processing using histogram (a) Original thermal image (b) Top 10% result. (c) Top 50% result. (d) Top 70% result.

B. Extrinsic calibration between thermal infrared camera and LiDAR

In order to use the information of two different sensors as a single fusion system, extrinsic calibration is required to combine the coordinate between the sensors into one coordinate system. There are several ways to integrate LiDAR and vision sensors. The conventional method is to obtain the matching point of two sensors using a board [2-5]. Recently, a method of extracting and fusing feature values has been studied [6]. There is no way to directly calibrate LiDAR and thermal infrared camera sensors. Zhang obtains the transformation matrix between the visual camera and LiDAR and then the transformation matrix between the thermal infrared camera and the visual camera. Finally, the extrinsic parameters between LiDAR and thermal infrared camera were calculated by multiplying the above two matrices [7].

In this paper, we propose algorithms that can be directly externally calibrated from thermal infrared cameras and LiDAR sensors. To extract the corresponding point of the two sensors, we first analyze the brightness distribution and saturation of the entire Thermal Image using a histogram, a technique for analyzing the frequency of image brightness values in thermal infrared cameras. As shown in Fig. 4, the 3D calibration targets is separated by a value corresponding to the histogram brightness value of the top $n\%$. n is the distribution ratio of the upper brightness values. Noise is eliminated from thermal image $t(x, y)$ using the histogram top $n\%$ value.

$$t(x,y) = \begin{Bmatrix} 0 & if\ \ t(x,y) > thresh \\ n & otherwise \end{Bmatrix} \tag{9}$$

From the extracted 3D calibration targets, points $tmax(t_x, t_y)$ (10) and $tmin(t_x, t_y)$ (11) are extracted by comparing pixel values in the thermal image $t(x, y)$.

$$tmax(x,y) = \begin{Bmatrix} max\ (t_x, t_y) & max(x,y) < t(x,y) \\ max(x,y) & otherwise \end{Bmatrix} \tag{10}$$

$$tmin(x,y) = \begin{Bmatrix} min\ (t_x, t_y) & min(x,y) > t(x,y) \\ min(x,y) & otherwise \end{Bmatrix} \tag{11}$$

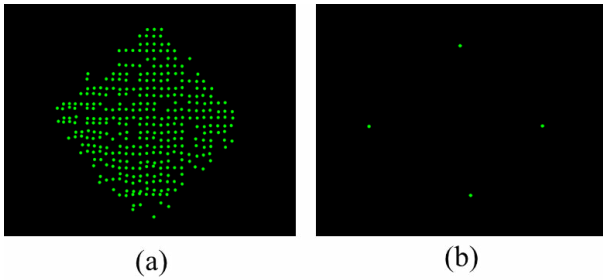

Fig. 5. LiDAR point extraction result (a)Extract 3D calibration targets. (b) Results of feature point extraction from the 3D calibration targets

As shown in Fig 5a, the 3D calibration targets is extracted from the LiDAR point cloud using the RANSAC plane fitting algorithm. Extract $lmax(t_x, t_y)$ (12) and $lmin(t_x, t_y)$ (13) points from the extracted 3D calibration targets $l(x, y, z)$.

$$lmax(x,y,z) = \begin{Bmatrix} max\ (l_x, l_y) & max(x,y) < l(x,y) \\ max(x,y) & otherwise \end{Bmatrix} \tag{12}$$

$$lmin(x,y,z) = \begin{Bmatrix} min\ (l_x, l_y) & min(x,y) > l(x,y) \\ min(x,y) & otherwise \end{Bmatrix} \tag{13}$$

All pairs of the corresponding points between the thermal infrared camera and the LIDAR were obtained to estimate the extrinsic calibration. To find the external parameters between the thermal infrared camera and LiDAR, we used OpenCV solvePnP algorithm (14)

$$s\begin{bmatrix} t_x \\ t_y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} l_x \\ l_y \\ l_z \\ 1 \end{bmatrix} \tag{14}$$

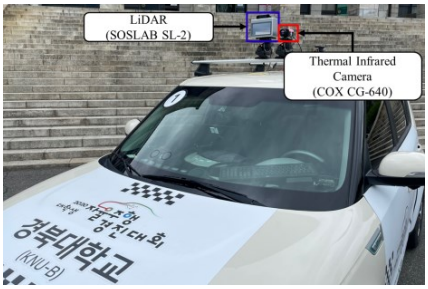

Fig. 6. Autonomous Vehicle Thermal Infrared Camera (COX CG-640), LiDAR (SOSLAB SL-2) demonstrating the proposed algorithm

III. Experimental Result

A. Experimental platform

As shown in Fig. 6, the sensor was composed of COX CG-640 thermal infrared camera (resolution: 640 x 480) and SOSLAB SL-2 LiDAR. The sensor is mounted on an autonomous vehicle. All algorithms run on desktop and use NVIDIA TITAN RTX D6 24GB, Intel® Core i9 9900k CPU @ 3.60GHz, 32GB RAM, Ubuntu 18.04, and ROS melodic.

Since the ground truth for the relative pose between the two sensors is difficult to obtain, the external parameters obtained cannot be compared with the ground truth. However, it is possible to evaluate the reprojection error by reprojecting from the LiDAR frame $\{L\}$ to the thermal infrared camera frame $\{T\}$. In addition, the thermal infrared camera frame $\{T\}$ can be visually evaluated by reconstructing it as a thermal point cloud.

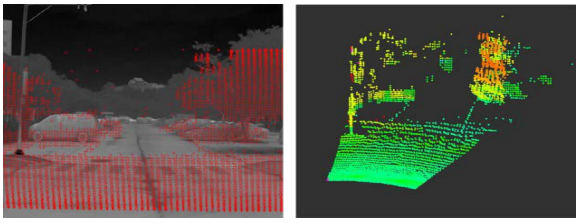

Fig. 7. (a) The result of projecting the LiDAR point as a thermal image. (b) Reconstructed thermal points

B. Thermal point cloud

To visually evaluate the external parameters between the thermal infrared camera and LiDAR, the LiDAR point P_l is converted into $P_t = \{X_t, Y_t, Z_t\}^t$ in the thermal infrared camera frame $\{T\}$ (15). P_t is projected onto the thermal image $p_t = (x_t,\ y_t)^T$, which is not distorted (16). The projected P_t is shown in Fig. 7a. A temperature value is assigned to the projected P_t. The projected point P_t can be reprojected to a point having a temperature value (17). The results are shown in Fig. 7b.

$$P_t = R_l^t P_l + T_{l;}^t \tag{15}$$

$$\begin{bmatrix} P_t \\ 1 \end{bmatrix} = \frac{1}{Z_t} K_t P_t \tag{16}$$

$$P_{thermal} = Z_t K_t^{-1} \begin{bmatrix} P_t \\ 1 \end{bmatrix} \tag{17}$$

C. Object detection using YOLOv4

For the object detection in an autonomous vehicle, research on learning and perception based on a lot of data is being activated. In this paper, when autonomous vehicles travel on the road, objects that can be judged as obstacles, vehicles and pedestrian are classified and detected. To detect the objects, YOLOv4 one of the Convolutional Neural Network (CNN) models, was applied [8]. YOLOv4 is a representative algorithm of one-stage detector and has the advantage of faster speed because localization and classification are performed at the same time. The object detection result is shown in Fig 8.

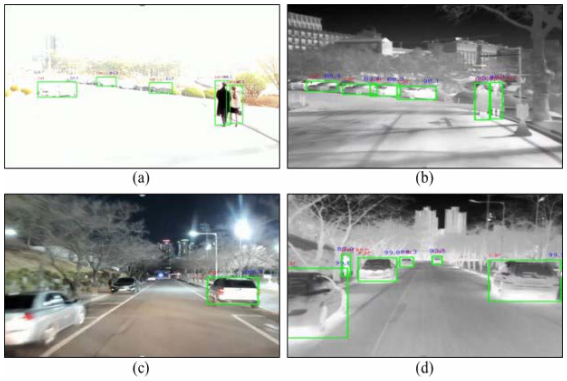

Fig. 8. Results of object detected in various lighting environments. (a) Car, pedestrian detection in the glare of sunlight. (b) Car, pedestrian detection in the dark light environment.

IV. Conclusion

In this paper, we propose an extrinsic calibration method to obtain external parameters between the thermal infrared camera and LiDAR. First, the proposed method is calibrated to obtain the intrinsic parameters of the thermal infrared camera. Second, in order to calibrate the thermal infrared camera and LiDAR, the histogram is analyzed from the 3D calibration targets in the thermal image, and noise is removed by applying an image threshold to the histogram. Points are extracted by comparing pixel values from the extracted 3D calibration targets. LiDAR extracts the 3D calibration targets through RANSAC plane fitting and compares the coordinate values of the points to extract the corresponding point. The two sensors are externally calibrated using the extracted corresponding points. Through experiments, the effectiveness of the above method has been proved. Through extrinsic calibration of thermal infrared cameras and LiDAR, autonomous vehicles can reliably detect objects even in situations where there is a lot of glare due to direct sunlight or headlights, and in environments where visibility is poor such as fog or smoke. In the future, 3D Object Detection system research will be conducted using thermal point cloud.

ACKNOWLEDGMENT

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] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1330-1334, Nov 2000.

[2] A. Dhall, K. Chelani, V. Radhakrishnan, K. Madhave Krishna, “LiDAR-Camera Calibration using 3D-3D Point correspondences," arXiv:1705.09785, pp. 1-19, May 2017.

[3] FM. Mizaei, DG. Kottas, SI. Roumeliotis, "3D LIDAR–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," The International Journal of Robotics Research, pp. 452-467, April 2012.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] P. An, T. Ma, K. Yu, B. Fang, J. Zhang, W. Fu, and J.Ma, "Geometric calibration for LiDAR-camera system fusing 3D-2D and 3D-3D point correspondences," Optics express, vol. 28, lssue 2, pp. 2122-2141, Jan 2020.

[5] A. Geiger, F. Moosmann, Ö. Car and B. Schuster, "Automatic Camera and Range Sensor Calibration using a single Shot," in 2012 IEEE International Conference on Robotics and Automation (ICRA), pp. 1-8, May 2021.

[6] X. Gong, Y. Lin and J. Liu, "3D LIDAR-Camera Extrinsic Calibration Using an Arbitrary Trihedron," Sensors,13(2), pp.1902-1918, Feb 2013.

[7] J. Zhang, P. Siritanawan, Y. Yue, C. Yang, M. Wen and D. Wang, "A Two-step Method for Extrinsic Calibration between a Sparse 3D LiDAR and a Thermal Camera," in 2018 International Conference on Control, Automation, Robotics and Vision (ICARCV), pp. 1039-1044, Dec 2018

[8] A. Bochkovskiy, CY. Wang, HYM. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv:2004.10934, pp. 1-17, Apr 2020