International Journal of Advanced Smart Convergence Vol.9 No.3 232-238 (2020)
http://dx.doi.org/10.7236/IJASC.2020.9.3.232

IJASC 20-3-30

Aerial Object Detection and Tracking based on Fusion of Vision and Lidar Sensors using Kalman Filter for UAV

Cheonman Park[1], Seongbong Lee[2], Hyeji Kim[3], Dongjin Lee[4*]

[1]*Master's student, Department of Aeronautical Systems Engineering, Hanseo University, Korea*
[2]*Undergraduate student, Department of Unmanned Aircraft Systems, Hanseo University, Korea*
[3]*Ph.D. student, Department of Aeronautical Systems Engineering, Hanseo University, Korea*
[4*]*Associate professor, Department of Unmanned Aircraft Systems, Hanseo University, Korea*
*{iwliy2025, koj331, rlagpwl668}@naver.com, *djlee@hanseo.ac.kr*

Abstract

In this paper, we study on aerial objects detection and position estimation algorithm for the safety of UAV that flight in BVLOS. We use the vision sensor and LiDAR to detect objects. We use YOLOv2 architecture based on CNN to detect objects on a 2D image. Additionally we use a clustering method to detect objects on point cloud data acquired from LiDAR. When a single sensor used, detection rate can be degraded in a specific situation depending on the characteristics of sensor. If the result of the detection algorithm using a single sensor is absent or false, we need to complement the detection accuracy. In order to complement the accuracy of detection algorithm based on a single sensor, we use the Kalman filter. And we fused the results of a single sensor to improve detection accuracy. We estimate the 3D position of the object using the pixel position of the object and distance measured to LiDAR. We verified the performance of proposed fusion algorithm by performing the simulation using the Gazebo simulator.

Keywords: *Vision Sensor, LiDAR, Sensor fusion, Kalman filter, Object detection, Position estimation*

1. Introduction

Demand for the development of collision avoidance technology for UAV(Unmanned Aerial Vehicles) is increasing for the safety of UAV that flight in BVLOS(Beyond Visual Line of Sight). Accordingly, studies such as object detection, collision prediction and collision avoidance technology using various sensors are being conducted. Among the necessary technologies to perform collision avoidance, object detection and position estimation algorithm are being studied using various sensors such as vision sensor and LiDAR. Jiwoon Lim et al. recognize workers in the robot system based on CNN(Convolutional Neural Network) technique, and estimate the position and velocity of the recognized workers using two vision sensors[1]. Yuanwei Wu et al. detect the object using the image background connectivity cue and track the object using the Kalman filter[2]. Kiuosumi Kidono et al. recognize pedestrians by applying SVM(Support Vector Machine) based on

Manuscript Received: August. 13, 2020 / Revised: August. 19, 2020 / Accepted: August. 26, 2020
Corresponding Author: djlee@hanseo.ac.kr
Tel: +82-41-671-6284
Associate Professor, Department of Unmanned Aircraft Systems, Hanseo University, Korea

3D point cloud data measured by LiDAR[3]. Bo Li et al. display the data measured by LiDAR on a 2D point map, and detect vehicles by inputting the data and bounding box obtained by clustering into FCN(Fully Convolutional Network)[4]. In the case of a detection algorithm using a single sensor, detection rate can be degraded in a specific situation depending on the characteristics of sensor. In order to compensate for this, algorithms to detect objects by fusing sensors are also being studied. Jongseo Lee at al. improve the object detection performance by fusion of the detection results using vision sensor and LiDAR, based on the late fusion method[5]. Cristiano Premebida at al. detect and track the object using LiDAR, and classify objects by fusing of LiDAR and vision sensor using Bayesian-sum decision rules[6].

In this paper, we propose aerial objects detection and position estimation algorithm using fusion of vision sensor and LiDAR. We use YOLOv2 architecture and a clustering algorithm for vision sensor and LiDAR each other. And we use the Kalman filter to improve the detection accuracy of algorithm using a single sensor, and fuse the supplemented estimates. Finally to estimate the 3D position of detected object, we use fused pixel position of the object and measured distance from LiDAR. In order to verify the detection and position estimation performance of the proposed algorithm, we perform the simulations using the Gazebo simulator.

2. Aerial Object Detection and Position Estimation

2.1 Aerial object detection using vision sensor

In this paper, YOLOv2 architecture based on CNN is used to detect aerial objects on an image acquired from the vision sensor[7]. We collect the learning data of objects that can be detected during flight such as birds and another UAVs. As a result of training using this learning data, we confirmed that objects are detected normally as shown in Figure 1. We assume that the center point of the detected object is the center point of detected rectangular area.

Figure 1. Object detection result using vision sensor

2.2 Aerial object detection using LiDAR

In order to improve the computational load of the detection algorithm based on LiDAR and prevent false detection caused by the ground, preprocessing of the measurement data was performed. Using the absolute altitude(h) shown in Figure 2, unnecessary ground information was removed as in Equation (1). And the information that outside the horizontal field of view(HFOV) of the vision sensor was removed as in Equation (2). In Equation (1), r is the measured distance, h_{extra} is the set value of the altitude for stable ground removal, and β is the angle between the measured channel from the axis perpendicular to the ground. In Equation (2), α_{min} and α_{max} is the minimum and maximum values of the azimuth set according to the HFOV of the vision sensor, and α is the measured azimuth.

$$r < \frac{h - h_{extra}}{\cos\beta} \tag{1}$$

$$\alpha_{min} < \alpha < \alpha_{max} \tag{2}$$

Note that set of the point cloud data that measuring on one cycle and removing ground is defined as P_t in Equation (3).

$$P_t = \{\boldsymbol{p}_{1,t}, \boldsymbol{p}_{2,t}, \cdots, \boldsymbol{p}_{n,t}\}, \qquad \boldsymbol{p}_{i,t} = [x_{i,t} \quad y_{i,t} \quad z_{i,t}]^T \tag{3}$$

Using P_t, we cluster the point cloud data. In Equation (4), $c_{i,t}$ is defined the set of clustered point cloud data. The condition for clustering is defined as in Equation (5). A point belongs to a group should satisfy that distance between other points in same group at least is smaller than threshold distance. This algorithm will be performed recursively until nothing is satisfied the condition in Equation (5). And we use the counts of clustered points to remove large or small objects such as building, trees and noise. In Equation (6), M_{min} and M_{max} need to decide by considering the operating environment and sensor resolution. If you need to detect a building, M_{min} and M_{max} should be large. We decide the position of the object as mean of each group as Equation (7). The Figure 3(a) represents concept of the clustering algorithm and Figure 3(b) shows the result of clustering algorithm with VLP16. In Figure 3(b), the zero value means detection fail.

$$C_t = \{c_{1,t}, c_{2,t}, \cdots, c_{n,t}\} \tag{4}$$

$$c_{i,t} = \{p \in P_t \mid \|p_m - p_n\| < d_{thresh},\ m \neq n\} \tag{5}$$

$$M_{min} < n(c_{i,t}) < M_{max}, \qquad M_{min} \geq 0 \ \ and \ \ M_{max} \geq 0 \tag{6}$$

$$\boldsymbol{\mu}_i = \frac{1}{n(c_{i,t})} \sum_{j=1}^{n(c_{i,t})} \boldsymbol{p}_{j,t} \tag{7}$$

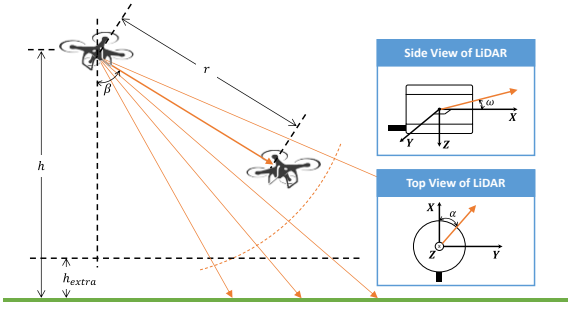
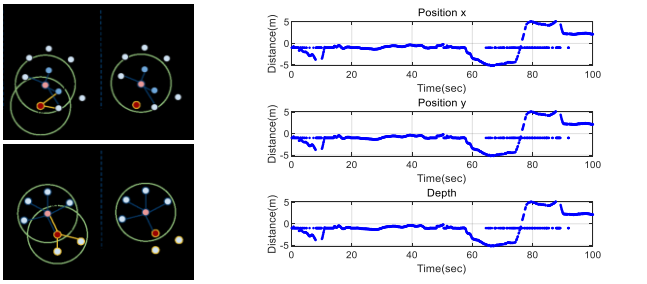


Figure 2. LiDAR system



(a) Clustering **(b) Center point estimation result using clustering**

Figure 3. Clustering results

2.3 Sensor fusion using Kalman filter

In this paper, an object detection algorithm based on sensor fusion is studied to improve the detection accuracy. To fuse two sensors, we use multiple estimated state fusion method. Each sensor measurement is used for Kalman filter[8] and we fuse the states of each Kalman filters. We use a linear fusion method using covariance[9]. The system model of the Kalman filter for estimating the state used a two-dimensional constant acceleration model for vision sensor as in Equations from (8) to (10), and three-dimensional constant acceleration model for LiDAR as in Equations from (11) to (13). In Equations, $\boldsymbol{x}$ is state variable, A, H, Q, R, P are parameters of Kalman filter, T is sampling time, α_{model}, α_{sensor} are variance values of each

system model and each sensor, and $\alpha_{initial}$ is initial variance value. Also in Equation (8) and (11), x_{pixel}, y_{pixel} are the position of x axis and y axis on an image, x, y are the position of y axis and z axis in LiDAR data, and d is depth value.

Model definition of vision sensor

$$\boldsymbol{x}_{2D} = \begin{bmatrix} x_{pixel} & \dot{x}_{pixel} & \ddot{x}_{pixel} & y_{pixel} & \dot{y}_{pixel} & \ddot{y}_{pixel} \end{bmatrix}^T \tag{8}$$

$$A_{2D} = \begin{bmatrix} T_{c,a} & 0_{3\times3} \\ 0_{3\times3} & T_{c,a} \end{bmatrix}, \quad T_{c,a} = \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}, \quad H_{2D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{9}$$

$$Q_{2D} = \alpha_{model} I_{6\times6}, \quad R_{2D} = \alpha_{sensor} I_{2\times2}, \quad P_{2D} = \alpha_{initial} I_{6\times6} \tag{10}$$

Model definition of LiDAR

$$\boldsymbol{x}_{3D} = \begin{bmatrix} x & \dot{x} & \ddot{x} & y & \dot{y} & \ddot{y} & d & \dot{d} & \ddot{d} \end{bmatrix}^T \tag{11}$$

$$A_{3D} = \begin{bmatrix} T_{c,a} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & T_{c,a} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & T_{c,a} \end{bmatrix}, \quad T_{c,a} = \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}, \quad H_{3D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{12}$$

$$Q_{3D} = \alpha_{model} I_{9\times9}, \quad R_{3D} = \alpha_{sensor} I_{3\times3}, \quad P_{3D} = \alpha_{initial} I_{9\times9} \tag{13}$$

The states estimated from each Kalman filter need to transform into the same dimension in order to fuse. So, we convert the 3D position information of LiDAR into the 2D pixel information as in Equations from (14) to (17). In Equation (15), x_{bias} and y_{bias} is the values to compensate for the mounting error between vision sensor and LiDAR, W is width of image, H is height of image, $HFOV$ is horizontal field of view, and $VFOV$ is vertical field of view. The Figure 4 represents that the overall data fusion procedure. Shown in Figure 4, each measurement is used for each constant acceleration Kalman filter. And we transform the 3D state into the 2D state in order to fuse each state of Kalman filter. States of each Kalman filter are fused linearly using covariance that means uncertainty of state of Kalman filter.

$$\boldsymbol{x}_{2D,L}(\hat{\boldsymbol{x}}_{3D}) = \begin{bmatrix} x_{2D,L} & \dot{x}_{2D,L} & \ddot{x}_{2D,L} & y_{2D,L} & \dot{y}_{2D,L} & \ddot{y}_{2D,L} \end{bmatrix}^T \tag{14}$$

$$x_{2D,L} = \frac{W}{2}\left(\frac{x}{d\tan\left(\frac{HFOV}{2}\right)} + 1\right) + x_{bias}, \quad y_{2D,L} = \frac{H}{2}\left(\frac{y}{d\tan\left(\frac{VFOV}{2}\right)} + 1\right) + y_{bias} \tag{15}$$

$$\dot{x}_{2D,L} = \frac{W}{2\tan\left(\frac{HFOV}{2}\right)}\left(\frac{\dot{x}d - x\dot{d}}{d^2}\right), \quad \dot{y}_{2D,L} = \frac{H}{2\tan\left(\frac{VFOV}{2}\right)}\left(\frac{\dot{y}d - y\dot{d}}{d^2}\right) \tag{16}$$

$$\ddot{x}_{2D,L} = \frac{W}{2\tan\left(\frac{HFOV}{2}\right)}\left(\frac{d(\ddot{x}d - \ddot{d}x) - 2\dot{d}(\dot{x}d - x\dot{d})}{d^3}\right), \quad \ddot{y}_{2D,L} = \frac{H}{2\tan\left(\frac{VFOV}{2}\right)}\left(\frac{d(\ddot{y}d - \ddot{d}y) - 2\dot{d}(\dot{y}d - y\dot{d})}{d^3}\right) \tag{17}$$

2.4 Position estimation for object tracking

The 3D position of the detected object is estimated using the depth of LiDAR and the center point on the image that is results of the fusion algorithm. Assuming that the camera intrinsic model K is known, and we can calculate the 3D position using camera geometry in Equation (18). In order to get the 3D position based NED coordinate, we use DCM(Direction Cosine Matrix, C) in Equation (19). We assume that the euler angle is measured by IMU sensor. In Equation (18), R is rotation matrix that converts to the image coordinate, and t is sampling time.

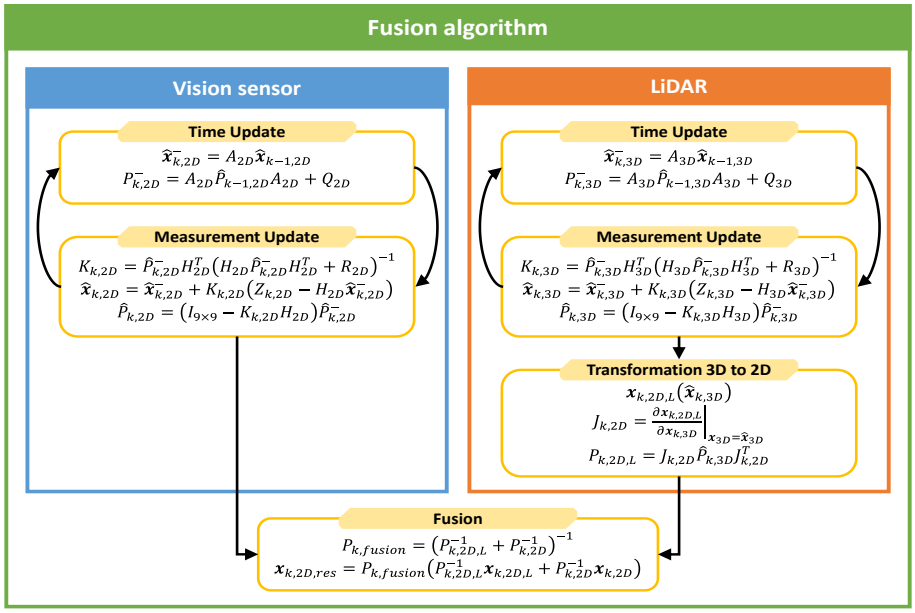


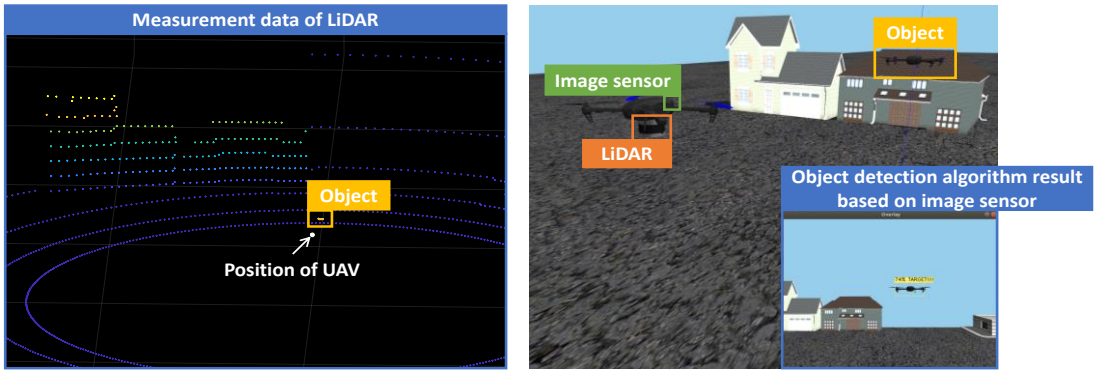
Figure 4. Procedure of sensor fusion algorithm using Kalman filter

$$\begin{bmatrix} x_{fused} \\ y_{fused} \\ 1 \end{bmatrix} = K[R|t]\begin{bmatrix} x_{body} \\ y_{body} \\ z_{body} \\ 1 \end{bmatrix} = K[R|t]X_{body} \tag{18}$$

$$C(\psi, \phi, \theta)X_{body} = X_{NED} \tag{19}$$

3. Simulation Results

In this paper, we simulated to verify the proposed algorithm. We use the Gazebo simulator for realistic scenario. Figure 5(a) shows measurement of LiDAR in simulation and Figure 5(b) shows the simulation environment. There are two UAVs, one is equipped with sensors to verify the algorithm and the other is a target. The specifications of the vision sensor and LiDAR used in simulation are set as shown in Table 1. The UAV and the object to be detect were set the Iris drone model.



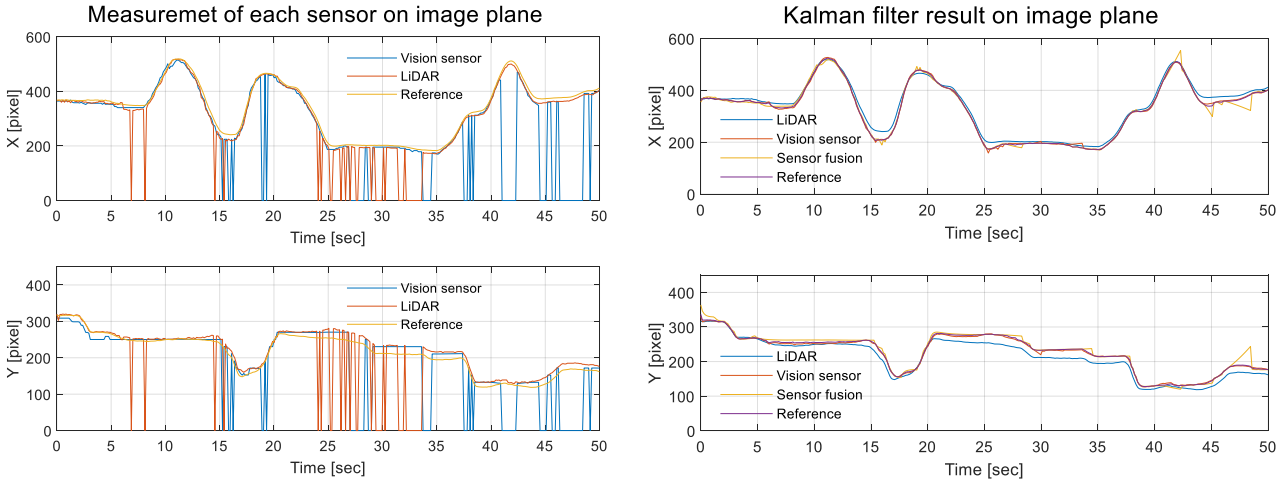
(a) Measurement data of LiDAR **(b) Simulation environment**

Figure 5. Simulation environment in Gazebo to verify performance of proposed algorithm

Table 1. Specification of vision sensor and LiDAR in Gazebo simulation

Vision sensor		LiDAR		
FOV (°)	Resolution (pixel)	FOV (°)	Range (m)	Channels
Horizontal : ±30 Vertical : ±22.5	Width : 640, Height : 480	Horizontal : 360, Vertical : ±15	Max 100	16

As a result of simulation, it was verified that the detection rate and accuracy of the sensor fusion based algorithm was improved than algorithms using each single sensor. Figure 6(a) shows measurement of each single sensor based algorithm. In Figure 6(a), the zero value means nothing was detected. Figure 6(b) shows each result of Kalman filters and sensor fusion. In certain circumstances, it could be seen that the detection rate of one sensor was degrader than one left, and sensor fusion based algorithm complement that. Figure 7 represents the position estimation result of proposed algorithm.



(a) Object detection result of each sensor **(b) Object detection result of each Kalman filter**

Figure 6. Object detection result of proposed algorithm

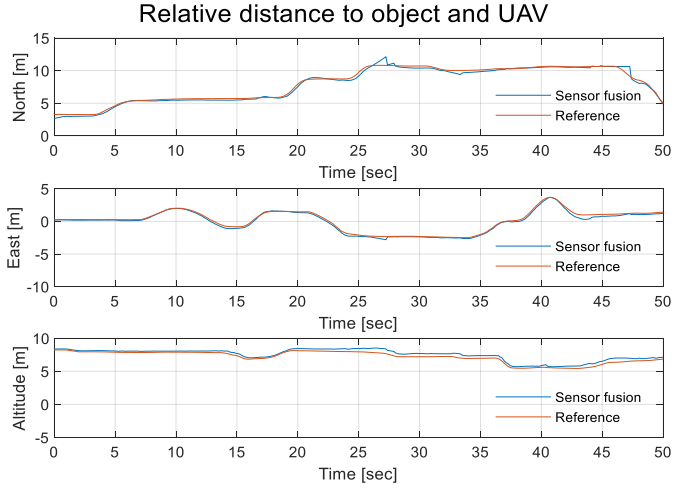


Figure 7. Position estimation result of proposed algorithm

4. Conclusions

In this paper, we studied on aerial objects detection and position estimation algorithm using sensor fusion for UAV. We verified performance of proposed algorithm through the simulation. There were some situations that LiDAR detection fail or vision sensor detection fail or both. When nothing was detected with vision sensor and LiDAR, the accuracy of proposed algorithm was degrade like others. But in whole of condition, detection and tracking performance were improved than single sensor based algorithm. If detection algorithm using each sensor is improved, we expect performance of proposed algorithm should be improved. So we will study to improve detection rate and accuracy of each single sensor, and to apply to collision avoidance this algorithm.

Acknowledgement

This work was supported by the Technology Innovation Program (20005015) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) in 2020.

References

[1] J.W. Lim, and S.S. Rhim, “Estimation of Human Position and Velocity in Collaborative Robot System Using Visual Object Detection Algorithm and Kalman Filter,” 2020 17th International Conference on Ubiquitous Robots (UR), June 2020. DOI: https://doi.org/10.1109/ur49135.2020.9144888.

[2] Y. Wu, Y. Sui and G. Wang, “Vision-Based Real-Time Aerial Object Localization and Tracking for UAV Sensing System,” IEEE Access, Vol. 5, pp. 23969-23978, October 2017. DOI: https://doi.org/10.1109/access.2017.2764419.

[3] K. Kidono, T. Miyasaka, A. Watanabe, T. Naito, and J. Miura, “Pedestrian recognition using high-definition LIDAR,” 2011 IEEE Intelligent Vehicles Symposium (IV), June 2011. DOI: http://dx.doi.org/10.1109/ivs.2011.5940433.

[4] B. Li, T. Zhang, and T. Xia, “Vehicle Detection from 3D Lidar Using Fully Convolutional Network,” Robotics: Science and Systems XII, August 2016. DOI: https://doi.org/10.15607/RSS.2016.XII.042.

[5] J.S. Lee, M.G. Kim, and H.I. Kim, “Camera and LiDAR Sensor Fusion for Improving Object Detection,” Journal of Broadcast Engineering(JBE), Vol. 24, No. 4, July 2019. DOI: https://dx.doi.org/10.5909/JBE.2019.24.4.580.

[6] C. Premebida, G. Monteiro, U. Nunes and P. Peixoto, “A Lidar and Vision-based Approach for Pedestrian and Vehicle Detection and Tracking,” 2007 IEEE Intelligent Transportation Systems Conference, pp. 1044-1049, October 2007. DOI: https://doi.org/10.1109/itsc.2007.4357637.

[7] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), July 2017. DOI: http://dx.doi.org/10.1109/cvpr.2017.690.

[8] N.A. Thacker and A.J. Lacey, *Tutorial: The Kalman Filter*, Tina Memo No. 1996-002, 1998

[9] D.G. Yoo, T.L. Song, and D.S. Kim, “Track-to-Track Information Fusion using 2D and 3D Radars,” Journal of Institute of Control, Robotics and Systems, Vol. 18, No. 9, pp. 863–870, September 2012. DOI: http://dx.doi.org/10.5302/j.icros.2012.18.9.863.