2020 20th International Conference on Control, Automation and Systems (ICCAS 2020)
Oct. 13∼16, 2020; BEXCO, Busan, Korea

Heterogeneous sensor fusion based omnidirectional object detection

Hyunjee Ryu[1] Inhwan Wee[2] Taeyeon Kim[3] David Hyunchul Shim[3]

[1]Division of Future Vehicle, KAIST,
Deajeon, 34141, Korea (localryu@kaist.ac.kr)
[2]KI for Robotics, KAIST,
Deajeon, 34141, Korea (finani@kaist.ac.kr)
[3]School of Electrical Engineering, KAIST,
Deajeon, 34141, Korea

Abstract: Nowadays, the importance of the object recognition of an aerial vehicle has increased, and many studies have been conducted. For Urban Aerial Mobility (UAM), it is important to recognize other vehicles, such as drones and birds, and avoid collisions when flying. In this paper, two sensors are fused to detect objects. It is used by fusing a camera sensor that can be used with light and low power, and a lidar sensor that has high near-field reliability and can know the location information of an object. Typically, Radar is used to recognize objects on airplanes but has been replaced by riders to conduct research on the drone platform. By using the features of the two sensors, the recognition rate of objects at short and long distances is increased, and reliability is increased through sensor redundancy. In addition, it was possible to drive in real-time on the embedded board through system optimization. Existing vehicles recognize other vehicles using Radar and communication. However, through the sensor fusion presented in this paper, it is possible to increase the object recognition rate in stand-alone situations.

Keywords: Sensor fusion, object detection, Deep-learning, real-time system, drone, UAM

1. INTRODUCTION

Recently, as interest in drone taxis has increased, the Urban Aerial Mobility (UAM) is increasing as well. So global companies are also entering UAM. In order for UAM to be commercialized, flight safety must be solved. Since UAM flies at low altitudes, collisions with aerial objects such as drones, birds, and balloons can occur. Therefore, UAM must be able to recognize and predict collisions to avoid collision with the aerial objects. Existing air vehicle recognition has mainly conducted by recognizing birds and drones using a camera sensor fixed on the ground, or by recognizing objects only in the direction of travel using one or two cameras for drones and airplanes. However, in the case of UAM and drone, collision avoidance with an aircraft is required in all directions. Existing studies did not address object recognition for all directions and did not guarantee real-time performance, so performance is insufficient for drones and UAMs. In this paper, we suggest a method and a system for object recognition in all directions using four cameras and a Light Detection and Ranging (LiDAR) sensor on a drone platform. Camera sensors are vulnerable to white-out or blackout, so using them alone may not be reliable. Moreover, the closer the own-aircraft is, the higher the recognition rate of the aircraft is required. To overcome these problems, we fuse four cameras and a lidar sensor. Through sensor fusion, we secure the redundancy of the sensor and increase the accuracy of object recognition in the near field. Through this system, we intend to increase the flight safety of UAM by recognizing aircraft.

2. SYSTEM

2.1 Hardware architecture

For detection, we use two types of sensors. In order to get the position, especially depth information, we use a LiDAR sensor instead of a Radar sensor which is usually used in aircraft. Another one is camera. The camera has a feature that can recognize objects at a relatively long distance with low power and lightweight.

Hardware system was designed and built for omni-directional object detection(Fig. 1). Gryphon Dynamics X8 for the drone's basic frame and the embedded board used Nvidia's Jetson Xavier. The LiDAR sensor used Velodyne 16-channel lite version and the camera sensors used 4 Ocams. In order to mount the 4 cameras and the LiDAR in the right direction, the mount was designed and manufactured with a 3D printer. Having a significant flight time, we constructed a distributed power system (Fig. 2) considering the power consumption.

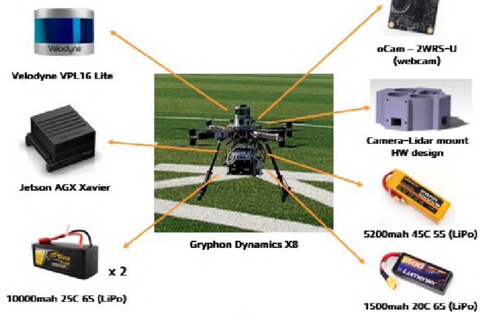

Fig. 1 Drone hardware architecture

In order to recognize omni-directionally with the cameras, the sensor part of the system was constructed using

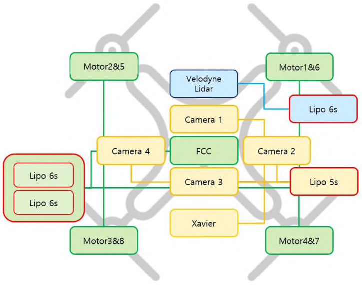

Fig. 2 Hardware power wiring diagram

4 webcams. In the Fig. 3, the lens is 110 degrees HFOV (Horizontal Field Of View). Therefore the hardware is configured to be able to see 360 degrees. The right images of Fig. 3 are from four cameras which shows the images coming in all directions.

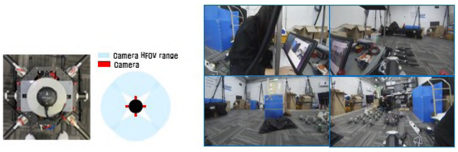

Fig. 3 4-camera system's FOV (left). the result (right)

2.2 Software pipeline

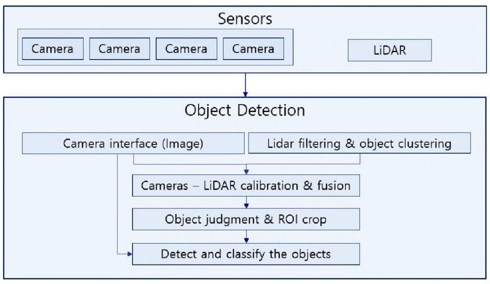

Fig. 4 Software pipeline

The software structure is shown in figure 4. Programs are developed on Robot Operating System(ROS) [1] and other well-established open-source software libraries. The object detection part extracts objects in parallel from the four cameras and a LiDAR sensor. If there is no LiDAR data, the images are directly used for object detection. Otherwise, if there is LiDAR data, the LiDAR data is filtered and clustered to extract something(objects). After then, using the processed information such as the position, select the matching direction's camera and image to detect the objects. Afterward, the images are used for object detection in a deep-learning based method.

3. OBJECT DETECTION

3.1 Object clustering and ROI cropping

Considering real-time calculations, we have constructed a system that recognizes objects using a LiDAR information. First, using the LiDAR point cloud, a Euclidean distance between two points is calculated, and if it is below a certain distance, it is regarded as the same cluster (nearest neighbors algorithm [2]-[3]). After that, calculate a center point of the cluster [3]. If the cluster size is smaller than a threshold, crop the ROI from the image using the LiDAR camera calibration matrix.

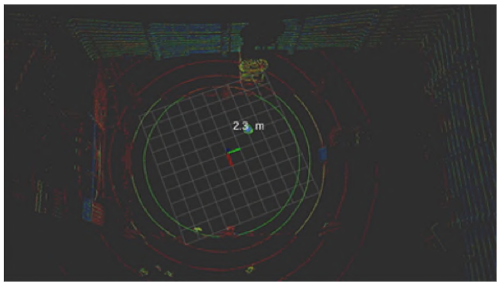

Fig. 5 Result of LiDAR data clustering and centroid

Fig. 6 Result of image cropping

3.2 Deep learning-based Object detection

Deep learning was used for object recognition and classification.

DJI Mavic Mini	350	DJI Phantom	200
USRG mini drone	200	BionicBird	200
F450	200	Real bird image	100
General drone image	100		

Fig. 7 Training dataset. For detect the small flight objects, we collected the data from several flight tests and downloaded the images from internet.

To create a training model, we made the dataset (figure 7.). The data were collected through several flight experiments. For the drone data, DJI Mavic mini and phantom, USRG mini drone, F450 drone and general drone images from Google were used. Moreover, for bird data, Bionicbird and general bird images from Google were used.

Architecture of detection model is used *MobileNetSSD* in figure 8. *MobileNetSSD* is a combination of convolutional neural network [4] and objects detection network [5].

When using only the camera sensor, the camera frames in four directions are used in inference in sequence. If an object is recognized in any plane, the camera frame in that direction is used until the object is no longer recognized.

Base model								
		TP	TN	FP	FN	Accuracy	avg. fps	avg.fps (detected)
Phantom	camera	203	354	0	274	0.670277	3.40	5.84
	fusion	349	334	0	168	0.802585	3.62	6.17
Mavic mini	camera	38	495	0	212	0.715436	2.51	4.96
	fusion	81	526	0	154	0.797635	2.67	5.92
TensorRT model								
		TP	TN	FP	FN	Accuracy	avg. fps	avg.fps (detected)
Phantom	camera	465	668	0	368	0.754830	6.10	13.20
	fusion	738	628	0	198	**0.873402**	6.31	**15.32**
Mavic mini	camera	82	588	0	223	0.750280	3.83	11.91
	fusion	119	603	0	140	0.837387	3.70	14.65

Table 1 Detection results. The results are compared in the Base model and TensorRT model. Test was conducted with two identical 'rosbag' data. The fusion method is more accurate than use (one type of sensor)cameras. Also, TensorRT model is better than the Base model in both fps and accuracy part.

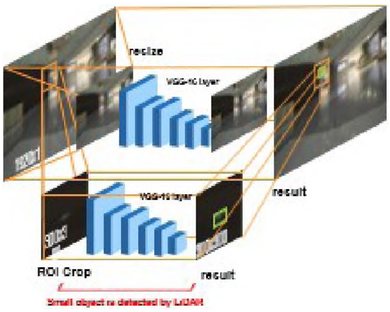

Fig. 8 Detection model architecture. Based on VGG-16 layer architecture.

If the object was not recognized in that frame, we again used the four-direction frame for the sequential inference. In this way, when an object exists or is recognized, fps may be higher than that recognized by inferring frames sequentially in four directions.

The recognition method using the LiDAR - camera sensor fusion was used to infer frames coming from the camera in the corresponding direction using the camera-LiDAR calibration matrix, if there are clustered objects after the LiDAR data preprocessing process. Even though the object was recognized through the LiDAR, the method of recognizing the camera was used for the accuracy of the detection. This method also helped to improve the fps when an object was present or recognized.

4. RESULTS

An omnidirectional objects(drones) recognition experiment was conducted through four cameras and a LiDAR. The figure below is the result of recognition. In the case of the first picture, the object is recognized only in the LiDAR sensor at a short distance (6.1m). The red circle in the figure presents the result of object detection by the LiDAR sensor. The second picture shows the object being recognized by both the LiDAR and the camera sensors at a short distance (11.2m). Since the object recognition results of the two types of sensors appear to be almost the same, the sensor data is well fused, and it shows that the sensor has redundancy through sensor fusion. The last one is when an object is recognized only by the camera sensor at a long distance (22.3m).

As a result of the experiment, it was found that the recognition rate of the lidar, lidar + camera sensor is high in the near field, and the recognition rate through the camera sensor increases as the distance increases. Like the proposed system model in the introduction, the reliability of recognition was increased through two sensors at a short distance, and a system that was recognized through a camera was developed at a distance.

Table 1 shows the recognition rate and average fps result by the algorithm proposed in this paper using a dataset obtained through the flight test. As can be seen from the table, it can be seen that the recognition method through sensor fusion is higher in detection rate and average fps than when using only camera sensors. In addition, the detection rate and fps could be dramatically improved by optimizing the TensorRT detection model in the embedded board Nvidia's Jetson Xavier.

As a result, a system that can recognize objects with an object recognition rate of 87% on the embedded board Nvidia's Jetson Xavier and detect objects in all directions in real time at an average of 15.32 fps when objects are recognized.

5. CONCLUSION

We developed a system for real-time object recognition in all directions of the drone through the fusion of camera and LiDAR sensor. By using heterogeneous sensors, the sensor redundancy was secured to improve recognition rate and recognition reliability. In addition,

Fig. 9 Object detection results. the result detected by LiDAR is shown in a red circle, and the result detected by camera is shown in a green box. LiDAR(left). LiDAR + camera(middle). camera(right)

by using different features of the camera and the LiDAR sensor, a system that increases the reliability of the near object recognition to the LiDAR sensor with high accuracy at a short distance, and increases the reliability of the far object recognition to the camera that is relatively good at recognizing far objects. Through this, it was possible to obtain the robust feature of recognition for the far/near object of the system. Moreover, through the optimization of the detection model, we have developed a system that runs in real-time on the embedded board. It is considered that the object detection system proposed in this paper is highly likely to be improved or used later. I think that this system can be used to make collision avoidance or collision determination later because it can find the position of the recognized object.

ACKNOWLEDGEMENT

This work was supported by Institute for Information communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.2019-0-01764, Development of AI-powered Autonomous Drone for Complex Indoor Environment)

REFERENCES

[1] Garage, Willow. ”ROS. org.” (2010).

[2] Rusu, Radu Bogdan. ”Semantic 3d object maps for everyday manipulation in human living environments.” KI-Künstliche Intelligenz 24, no. 4 pp 345-348. 2010.

[3] Kato, Shinpei, Shota Tokunaga, Yuya Maruyama, Seiya Maeda, Manato Hirabayashi, Yuki Kitsukawa, Abraham Monrroy, Tomohito Ando, Yusuke Fujii, and Takuya Azumi. ”Autoware on board: Enabling autonomous vehicles with embedded systems.” *In 2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*, pp. 287-296. IEEE, 2018.

[4] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. ”Mobilenets: Efficient convolutional neural networks for mobile vision applications.” arXiv preprint arXiv:1704.04861. 2017.

[5] Huang, Jonathan, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer et al. ”Speed/accuracy trade-offs for modern convolutional object detectors.” *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7310-7311. 2017.