Real-Time Monocular Vision-Based UAV Obstacle Detection and Collision Avoidance in GPS-Denied Outdoor Environments Using CNN MobileNet-SSD

Daniel S. Levkovits-Scherer[1], Israel Cruz-Vega[2](✉), and José Martinez-Carranza[3,4]

[1] Electronics and Circuits Department, Universidad Simon Bolivar, Caracas, Venezuela
daniellevkovits@gmail.com

[2] Electronics Department, Instituto Nacional de Astrofisica, Optica y Electronica, San Andrés Cholula, Puebla, Mexico
icruzv@inaoep.mx

[3] Computer Science Department, Instituto Nacional de Astrofisica, Optica y Electronica, San Andrés Cholula, Puebla, Mexico
carranza@inaoep.mx

[4] University of Bristol, Bristol BS8 1UB, UK

Abstract. In this paper, we propose a monocular vision-based system that uses a MobileNet-SSD CNN for obstacle detection and collision avoidance in GPS-denied outdoor environments. This framework consists of two processes carried out simultaneously in a frame-to-frame basis: (1) an obstacle detector and classifier using a lightweight convolutional neural network with a UAV monocular onboard camera for real-time mobile systems; (2) a collision avoidance algorithm with a proportional controller responsible for the autonomous flight in GPS-denied outdoor environments. However, because object detection and classification are computationally intensive tasks, the processing is carried out off-board on a ground control station that receives online imagery and data of the UAV during the autonomous flight. The novel aspects in this work are related to the capacity of the system to detect and avoid obstacles in real-time with computationally low range hardware without GPU. We exploit public datasets meant for other purposes and carefully selected images to build a new lightweight dataset to train the CNN. Further, the output imagery data is used by a proportional controller that communicates back to the vehicle to evaluate a possible obstacle avoidance trajectory and execute it if necessary. We carried out evaluations and flights in real scenarios with multiple obstacles such as vehicles, people, bicycles, and trees for autonomous flights in GPS-denied outdoor environments with promising results.

L. Martínez-Villaseñor et al. (Eds.): MICAI 2019, LNAI 11835, pp. 613–621, 2019.
https://doi.org/10.1007/978-3-030-33749-0_49

Keywords: Autonomous flight · CNN · Collision avoidance · GPS-denied outdoor environments · MobileNet-SSD · Obstacle detection.

1 Introduction

Unmanned Aerial Vehicles (UAVs), have had a peak since last decade for commercial and scientific use, driving the reduction of costs and miniaturization, thus achieving to increase the number of applications for which they are currently used. Some of the applications range from recreational use, manipulation of objects, transport of lightweight loads to deployment in outdoor areas of difficult access for exploration missions and rescues and victims identification in natural disasters.

The effort to create and improve drone trajectory planning and control strategies has become a scientific and research trend. As well as using drones integrated onboard cameras for applications beyond photography, such as real-time image processing and object detection. As a result of the integration of these, the UAVs began to be used to carry out different tasks autonomously.

Nowadays, drones used for autonomous navigation make use of programmed trajectories, relying entirely on sensors such as GPS and ultrasound to avoid collisions, which have a high consumption of current and memory, reducing flight time and processing speed and thus resulting in low performance for assigned tasks. Another main issue to be addressed is that of drone localization given the partial or total loss of GPS signal.

Motivated by the above, in this work, we address the problem of autonomous flight in outdoor environments with limited or denied GPS signal by using a low-end development UAV and base control station in an efficient way. As well as a monocular onboard camera to process the environment to classify, detect and locate objects such as trees, people, bicycles and vehicles in real-time and process the information for a posterior collision avoidance without any other positioning technique or technology. Besides, we are motivated by the idea of achieving autonomous flight outdoors by using the least set of sensors, this is, a monocular camera and an altimeter, which is attractive in terms of energy consumption efficiency, an incentive for the development of micro aerial unmanned vehicles.

The above calls for a method that enables the drone to autonomously decide whether a detected object is far or close depending on the class and the obstacle-image relation is shown in Fig. 1. Thus we propose a two-step methodology for the autonomous flight where only a monocular onboard camera is used to carry out the detection. The process involves two steps carried out on a frame-to-frame basis. First, the image captured is processed by a Convolutional Neural Network (MobileNet-SSD CNN) [7], whose output will be one of the four classes: People, Bicycle, Tree, Vehicle; and coordinates of each object in the image. A second step will determine whether the object is considered as a proximate collision and avoid it using a reactive control algorithm.

To achieve the above, the CNN architecture [11] has been trained with a modified PASCAL VOC 2007 dataset and a hand made dataset, using only pertinent classes for training the CNN. All the images came from real environments, yet the network is capable of generalizing the characteristics for classification and detection for simulated obstacles resulting in promising results in simulated outdoor environments as well.

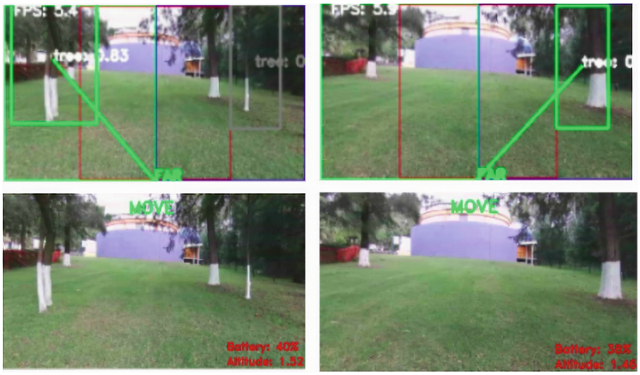

Fig. 1. We present a methodology to achieve autonomous flight outdoors with a drone equipped with a monocular camera. We use a deep learning approach [6] to obtain the class and coordinates of the image and thus the proximity and location on a frame-to-frame basis. All this by passing the UAV camera image through a MobileNet-SSD Convolutional Neural Network.

The rest of this paper is organized as follows: Sect. 2 describes relevant related work; Sect. 3 describes our proposed methodology; Sect. 4 describes our experimental framework; finally, our conclusions are discussed in Sect. 5.

2 Related Work

Currently, most of the work in the area is related to autonomous flights using GPS for greater accuracy, which results in higher battery consumption. For this reason, more efficient alternatives are sought that do not involve the use of GPS. When looking for other options in the literature, we found work focused on the use of LSD-SLAM for autonomous navigation [12], as well as Visual SLAM [1], without using GPS in uncontrolled environments [8]. One of the disadvantages presented in [12] is that the mapping cannot be done in real-time, since the need to pause to map the environment and the visual field, being limited by the number of points the system can map, and repeating this process periodically as you go along the environment being mapped. This method is useful for indoor environments with large objects density, as shown in Fig. 2.

Another precise method, but more useful in controlled environments is the VICON system [9], which locates the vehicle with millimetric precision in a 3D space, making use of markers avoiding GPS usage. This system is efficient and robust for evaluating control algorithms and drone specific behaviors, by investigating its effect when moving close to surfaces, as well as to check the stability of air vehicles in general. By being able to map with great precision, this system has an elevated cost in comparison to the vehicle used for the work. This technique requires several infrared cameras for 3D movement detection, and for the measurement to be precise, it is necessary to place markers on the UAV.

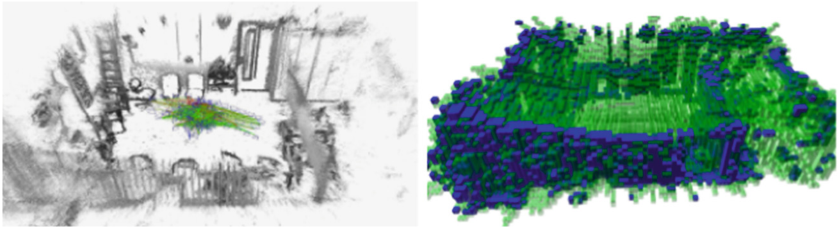

Fig. 2. Semidense reconstruction and 3D map of indoor environment mapped with LSD-SLAM [12]

Finally, all of the methods above explained need high processing capacity to function correctly in a real-time autonomous flight [10]. Because of these, we proposed a technique handled with low-cost equipment and processing units, without the need for dedicated GPUs or stereo cameras. We are using only an integrated monocular camera with a low-resolution image processing and without GPS sensor usage, all of the above with a deep learning approach [6]. We employ a MobileNet-SSD Convolutional Neural Network to classify [3], detect and locate possible collision targets and avoid them if necessary using a simple linear feedback reactive control [5].

3 Methodology

Our approach is based on three main components: (1) MobileNet-SSD CNN architecture as an object detector for a monocular onboard camera system; (2) proximity estimation in a single image using bounding box coordinates and aspect ratio; (3) and a Proportional controller.

In this work, we use a quadrotor vehicle with a monocular camera onboard. However, the RGB image acquired with the onboard camera is passed to the Convolutional Neural Network escalade to 320×180 pixels, thus generating an output of the object class and bounding box coordinates of the above, thus obtaining a proximity estimate and an object-image ratio, information that is used by the controller to generate a frame-to-frame basis flight plan. We implemented a P controller to control roll, pitch, yaw, and altitude velocity and thus to achieve drones motion control for an efficient autonomous flight with collision avoidance capabilities without using GPS or other techniques.

The controller error depends exclusively on each speed coordinate. Pitch error depends on the proximity of an object to the UAV, roll velocity error depends on the desired time to avoid the obstacle; yaw error depends on drone deviation angle from taking off position, and altitude velocity error depends on the difference between a desired fixed Z coordinate and actual altitude.

3.1 Simulated and Real Outdoor Environments

The classification and detection of a determined object in images are addressed using a methodology that exploits visual information by processing image pixels from publicly available datasets imagery. For our purpose, we processed images from the Pascal VOC 2007 dataset and a handmade dataset to extract these characteristics maps from both RGB images and bounding box ground truth. The idea behind this approach is to take advantage of the ground truth data available in the dataset imagery, extracting enough characteristics to train the CNN.

The simulated environment aims to establish the MobileNet-SSD capacity of generalization when training with a purely realistic images dataset of people, trees, bicycles, and vehicles, see Fig. 3.

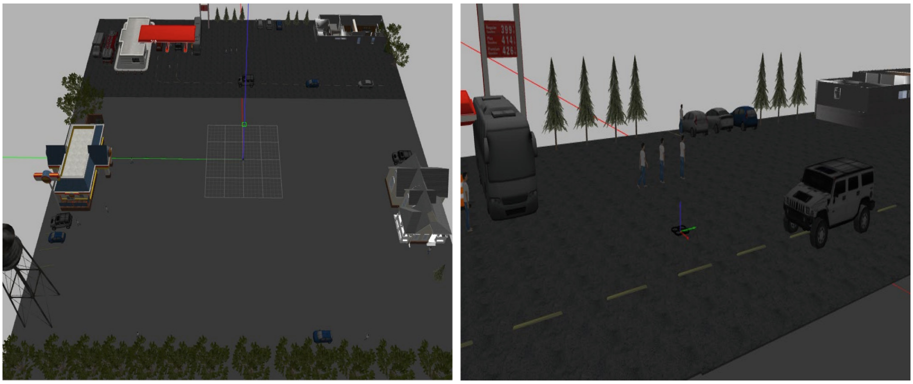

Fig. 3. Simulated outdoor environment for evaluation of our approach in Gazebo 7.

Motivated on the success of the approach above, we modified the algorithms and the communication system to make autonomous flight trials with real UAV and outdoor scenarios. For this, we modified all experimental parameters used for proximity sensing and obstacle avoidance since image- object ratios change drastically compared to simulated scenarios.

3.2 MobileNet-SSD

The MobileNet-SSD network aims to classify, detect and locate objects in an image for a moving robot without the necessity to stop moving to process the environment, hence his lightweight architecture is capable of extracting the necessary data with only one pass through the network. To reduce training time we use Transfer Learning technique [2], this CNN has a pre-trained base net [4] (MobileNet) which characteristics map is used by the SSD (Single Shot Multibox Detector) to improve learning and generalization and determine best adjustable bounding boxes for each object detected.

4 Experiments and Results

In this section, we describe the evaluation results, as well as the experiments realized in the Gazebo simulation world and the tests, accomplished in outdoor environments with one and multiple obstacles. Finally, we compare both tests to demonstrate the effectiveness of the system. To evaluate the performance of our detection capability, we used a dataset containing 6600 images of the four above mentioned classes with a 1-1 proportion for each one. The training dataset contains all the information about ground truth boxes and classes of the objects in each image. We compared the evaluation results of our simulated outdoor environment detector with the real outdoor environment results, carrying out several dozens of flights for each class in both simulation and real environments.

4.1 Simulation Outdoor Environment

For the first one with only an obstacle, the autonomous flight detector achieved 75% for people detection and avoidance, 80% for trees and 100% effectiveness for vehicles. These results reaffirm the capability of the MobileNet-SSD CNN of characteristics generalization due to the network only trained with real environment imagery, see Figs. 4 and 5.

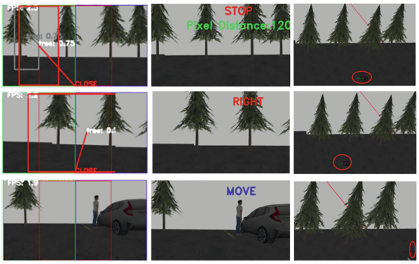

Fig. 4. Tree detection, and collision avoidance during autonomous flight.

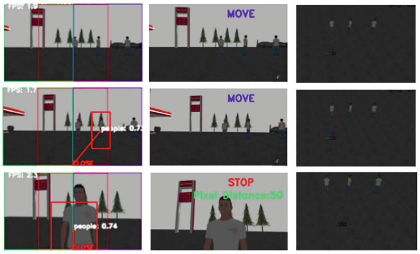

Fig. 5. People detection, and collision avoidance during autonomous flight.

4.2 Real Outdoor Environment

The second experiment was developed in a real outdoor rural environment where flights were done in various scenarios with different obstacles. Tree class was used for multiple obstacle autonomous flight tests to ensure the system capability to avoid obstacles in a continuous way, see Figs. 6 and 7.

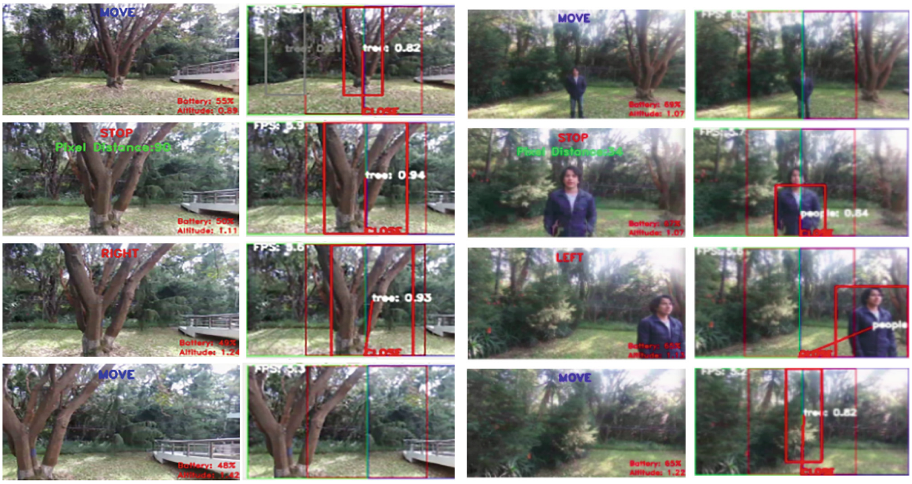

Fig. 6. Tree and people detection and collision avoidance in a real environment with one obstacle during autonomous flight

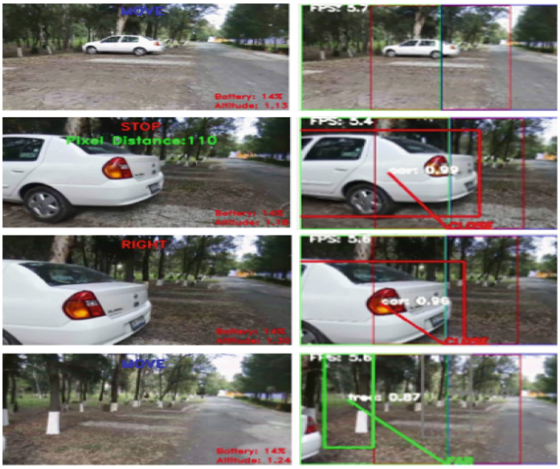

Fig. 7. Vehicle detection and collision avoidance in a real environment with one obstacle during autonomous flight

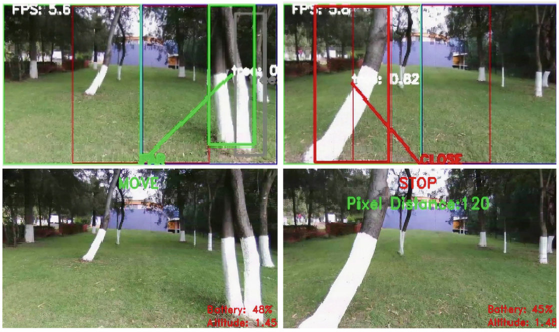

Fig. 8. Multiple tree detection and collision avoidance in a real environment with one obstacle during autonomous flight

The results when carrying out real environment autonomous flights had an improvement of 15% for people and tree classes and a decretion of 5% for vehicles. This improvement is due to the CNN training with real imagery datasets. And the decretion in detection for vehicle class is due to wind and illumination condition in real environments causing the system to collide with the obstacles.

Finally, to demonstrate the capability of the system to carry out autonomous flights with more than one obstacle to evade, as shown in Fig. 8, we made several flights with 2, 3 and 5 objects and the percentage of success went down to 90% for two obstacles and 85% for five obstacles.

5 Conclusions

This paper presents an autonomous flight system for a UAV in environments in GPS-denied areas, using only a monocular RGB camera, an integrated ultrasound sensor, and low-end equipment. A deep learning approach was used, explicitly applying convolutional neural networks for classification and detection of the environment in real time. Not only was it possible to evaluate the feasibility of a reactive autonomous flight policy as initially proposed, but also carry it out in a real system by obtaining optimal results.

The system approach is to use the MobileNet-SSD CNN, which processes the input image, i.e., the classification, detection, and location of the object in the image. For this the network is trained with four classes: vehicles, people, trees and bicycles, which are common in areas with limited or no access to GPS, such as rural areas.

The results presented demonstrate that the proposed method of autonomous navigation provides optimal results; the UAV can effectively execute its flight plan through the different objects offered with rates rounding 90% success of flights performed.

Acknowledgments. We thank the Instituto Nacional de Astrofisica, Optica y Electronica of Mexico (INAOE) for the laboratories and equipment borrowed and Eng.

Roberto Munguia-Silva for the continuous support and collaboration with time and knowledge throughout this project development.

References

1. Blösch, M., Weiss, S., Scaramuzza, D., Siegwart, R.: Vision based MAV navigation in unknown and unstructured environments. In: 2010 IEEE International Conference on Robotics and Automation, pp. 21–28. IEEE (2010)
2. Brownlee, J.: A gentle introduction to transfer learning for deep learning (2017). https://machinelearningmastery.com/transfer-learning-for-deep-learning/
3. Forson, E.: Understanding SSD multibox-real-time object detection in deep learning (2017). https://towardsdatascience.com/understanding-ssd-multibox-real-time-object-detection-in-deep-learning-495ef744fab
4. Howard, A.G., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
5. Kada, B., Ghazzawi, Y.: Robust PID controller design for an UAV flight control system. In: Proceedings of the World Congress on Engineering and Computer Science, vol. 2 (2011)
6. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436 (2015)
7. Li, Y., Huang, H., Xie, Q., Yao, L., Chen, Q.: Research on a surface defect detection algorithm based on mobilenet-SSD. Appl. Sci. **8**(9), 1678 (2018)
8. Mancini, M., Costante, G., Valigi, P., Ciarfuglia, T.A.: J-mod 2: joint monocular obstacle detection and depth estimation. IEEE Robot. Autom. Lett. **3**(3), 1490–1497 (2018)
9. Mashood, A., Mohammed, M., Abdulwahab, M., Abdulwahab, S., Noura, H.: A hardware setup for formation flight of UAVs using motion tracking system. In: 2015 10th International Symposium on Mechatronics and its Applications (ISMA), pp. 1–6. IEEE (2015)
10. Nemati, A., et al.: Autonomous navigation of UAV through GPS-denied indoor environment with obstacles. In: AIAA SciTech, pp. 5–9 (2015)
11. Saha, S.: A comprehensive guide to convolutional neural networks - the eli5 way (2018). https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
12. von Stumberg, L., Usenko, V., Engel, J., Stückler, J., Cremers, D.: Autonomous exploration with a low-cost quadrocopter using semi-dense monocular slam. arXiv preprint arXiv:1609.07835 (2016)