Using deep learning vision technology to control drone dynamic motions

1st Hyeongbin Park
dept. Information Convergence
Kwangwoon University
Seoul, Republic of Korea
mtanger@kw.ac.kr

2nd Yujin Lee
dept. Software
Kwangwoon University
Seoul, Republic of Korea
yj5878@kw.ac.kr

3rd Soeun Lee
dept. Computer Science and Statistics
Jeju National University
Jeju, Republic of Korea
thdms8477@jejunu.ac.kr

4th Seunghwan Kim
dept. Big Data Convergence
Jeju National University
Jeju, Republic of Korea
franz0602@stu.jejunu.ac.kr

5th Xinning Bai
dept. Computer and Information Technology
Purdue University
West Lafayette, IN, United States
bai110@purdue.edu

6th Jake Hazelton
dept. Computer and Information Technology
Purdue University
West Lafayette, IN, United States
jhazelt@purdue.edu

***Abstract*—With the rapid development of technology, UAVs have become increasingly popular among industries; however, they are not widely used by the general public due to the high cost of prior professional training and resources for controlling drones. In order to break those barriers, this document provides an efficient way to control dynamic drone movements. The two core accomplishments of the project are: (1) using deep learning trained on hand gesture images and (2) machine learning trained on 3D coordinate systems. In contrast to previous approaches, additional sensors and devices are replaced by AI methods, and the optimal datasets and deep learning algorithms are explored. After the safety testing of the entire system, further vision-based methods would be discussed.**

***Index Terms*—deep learning, machine learning, computer vision, UAVs, drone**

I. Introduction

The interest in Unmanned Aerial Vehicles, UAVs are gradually increased across a various industries in recent years. Though there is a high demand for UAVs in industries, drones are still not widely accessible to the public. In reality, it needs professional training time to control the drone and a lot of resources costly to be accustomed to the drone controller. Therefore, this project aims to develop a new human-drone interaction system to control drones easily by gestures without spending money and time on prior training and the conventional drone remote controller.

Thus, this document describes the solution to the problem of controller using computer vision-based deep learning methods. Google Mediapipe was used to detect the 3D coordinates systems of hand, pose, and face from the mobile camera. The classifier developed in this project will receive the 3D coordinate datasets detected by Mediapipe. Then, the model was trained to recognize the user’s gestures and motions based on datasets. The more accurate the classifier is, the more stable movement of the drone is controlled. Finally, this method is proposed to benefit both the public and the industries. It is expected to draw the attention of the public to using drones and help industries cut the extra cost of implementing drone in their business.

II. Literature Review

In most experiments, a wearable device or additional sensor, such as Leap Motion Controller was used. A Leap Motion Controller recognizes the user’s hand and automatically converts it into skeleton data. Though, this method requires a separate device other than the camera that is basically used to protect the hardware to keep the system performance.

The amount of research studying gesture-controlled drones has significantly increased in recent years. Most experiments [1] - [3] use wearable devices or additional sensors, such as a Leap motion controller to control the drone. The Leap motion controller aims to recognize the user's hand and automatically convert the hand's images into skeleton data for later use. Though the sensor works well converting data to control drones, the maintenance cost of the extra device other than the camera that is basically used is high. It is also inevitable to protect those devices from environmental factors, including wind speed and extreme weathers, for the whole system performance.

A few research papers have suggested using artificial intelligence (AI) and big data analysis to replace the sensor. In computer vision field, there is an object detection subset and there are several AI APIs developed for recognizing a specific object. Among them, Mediapipe from Google is the lightweight, fast, and flexible machine learning API trained on massive and various Google image datasets. Authors [4] implemented Mediapipe in their study to get hand key point datasets and built the classifier model to recognize eight customized gestures. This means that extra devices could be replaced by machine learning easily. In [5], authors also implemented a similar API, OpenPose, developed by Carnegie Mellon University team. OpenPose is a deep learning framework to estimate 2D coordinate key points from the image and used to build the low source framework for human-drone interaction in the paper. Nevertheless, there are some drawbacks such as the system only functions on the trained gestures, only the right hand, not the left hand or both hands with complex dynamic movements, and needs specific support using vision technology. This is regarded as bias that occurs in the process of collecting datasets and generalization, or overfitting, that model did not learn while training.

Therefore, several factors affecting the entire system performance mentioned in the most recent documents could be summarized as follow. Hardware factors include the capability of an embedded GPU system [6], the number of drones being controlled at once [5], and the estimated distance between the user and the drone [7]. Software factors include unexpected datasets in outdoor conditions, bias on trained datasets, and architecture of the classifier model [8]. This paper focuses on the challenges from software factors and gives computer vision solutions to overcome that.

III. Methodology

IV. Implementation

V. Conclusion

Acknowledgment

References

[1] J. Jun, B. Zhao, P. Zhang, F. Jia and P. Fang, "Interaction with UAV Formation based on Gesture Recognition with Leap Motion," 2021 33rd Chinese Control and Decision Conference (CCDC), 2021, pp. 3275-3280, doi: 10.1109/CCDC52312.2021.9602101.

[2] C. -C. Tsai, C. -C. Kuo and Y. -L. Chen, "3D Hand Gesture Recognition for Drone Control in Unity," 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), 2020, pp. 985-988, doi: 10.1109/CASE48305.2020.9216807.

[3] A. Menshchikov et al., "Data-Driven Body-Machine Interface for Drone Intuitive Control through Voice and Gestures," IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, 2019, pp. 5602-5609, doi: 10.1109/IECON.2019.8926635.

[4] V. K. Nguyen and R. Alba-Flores, "UAVs Control Using 3D Hand Keypoint Gestures," SoutheastCon 2022, 2022, pp. 140-144, doi: 10.1109/SoutheastCon48659.2022.9764030.

[5] M. Bhat, G. Mahto, S. Kesaria, V. Femandes and K. Arya, "Real-time gesture control UAV with a low resource framework," 2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA), 2021, pp. 19-24, doi: 10.1109/IRIA53009.2021.9588681.

[6] A. Menshchikov et al., "Data-Driven Body-Machine Interface for Drone Intuitive Control through Voice and Gestures," IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, 2019, pp. 5602-5609, doi: 10.1109/IECON.2019.8926635.

[7] F. Naseer, G. Ullah, M. A. Siddiqui, M. Jawad Khan, K. -S. Hong and N. Naseer, "Deep Learning-Based Unmanned Aerial Vehicle Control with Hand Gesture and Computer Vision," 2022 13th Asian Control Conference (ASCC), 2022, pp. 1-6, doi: 10.23919/ASCC56756.2022.9828347.

[8] B. Hu and J. Wang, "Deep Learning Based Hand Gesture Recognition and UAV Flight Controls," 2018 24th International Conference on Automation and Computing (ICAC), 2018, pp. 1-6, doi: 10.23919/IConAC.2018.8748953.