110

Real-time Video Streams of Hand Gestures to Control Unmanned Aerial Systems (UAVs)

Van Khoi Nguyen
Department of Electrical and Computer Engineering

Faculty Mentor: Dr. Rocio Alba-Flores

INTRODUCTION

Over past few years, unmanned aircraft vehicles (UAVs) have been becoming more and more popular for various purposes such as surveillance, automated industry, robotics, vehicle guidance, traffic monitoring and control system. It is very important to have multiple methods of UAVs controlling to fit in UAVs usages. The goal of this work is to develop a technique to control an UAV using 8 different hand gestures. To achieve that, a hand key point detection algorithm is developed to detect 21 key points of the hand. Then, those key points are used as the input to an intelligent system based on Convolutional Neural Networks (CNN) that is able to classify the hand gestures. After archiving hand gesture classifications, unique command for UAVs control is assigned to each hand gesture; and a programming is built in Python to send those commands to an UAV for operation.

Figure 1. DJI Ryze Tello UAV

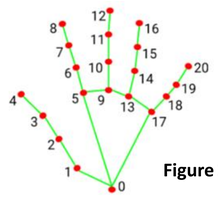

Figure 2. Hand Keypoints [1]

METHODODOLOGY

The main steps that were performed to achieve the goal are illustrated in fig. 3, and are as follows.

1) Images from the UAV camera are captured, and the system waits until a hand in the image is detected
2) Once a hand has been detected, a hand key point model is used to locate 21, 3D hand key point coordinates
3) Then, the meaning of the gesture in the input image is determined through the positional relationships between the key nodes
4) The hand gesture is classified using artificial neural networks
5) The result of the hand gesture classification is then used to generate control signals for the drone.

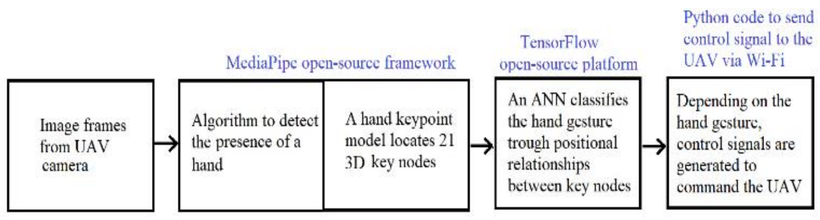

Figure 3: Block Diagram of UAV Control System

In this project, 2400 data of Hand key points were collected, preprocessed, and feed into an ANN algorithm for classification of eight different Hand Gestures. After the classification process, a Python programming was built to host a connection between UAV and a laptop. Python programming is used to set a unique commands for each classification sending to the UAV. The detail of Hand Gestures and its commands is shown in fig. 4.

Classification	**0**	**1**	**2**	**3**
Hand Motion	Hand Open	Hand Closed	Pointer	Thumb Up
UAV Motion	Fly Forward	Fly Backward	No Motion	Fly up
Example				
Classification	**4**	**5**	**6**	**7**
Hand Motion	Thumb Down	Thumb Left	Thumb Right	OK
UAV Motion	Fly Down	Fly Left	Fly Right	Land
Example				

Figure 4. Hand Gesture for UAV Control

There are totally 8 different hand gestures: hand open, hand close, pointer, thumb up, thumb down, thumb left, thumb right, and the OK sign. Each hand gesture is represented for an UAV command.

METHODOLOGY (cont'd.)

To detect the hand pose we based our project on the MediaPipe framework to make the process of hand key point detection faster and more precise. MediaPipe [1] is a graph-based framework for building multimodal (video, audio, and sensors) applied machine learning pipelines. This approach consists in using convolutional neural networks to obtain rough estimates of hand key points. This detection method is used for single images and it runs in real time on RGB images, and it has an accuracy comparable to methods that uses depth sensors.

There are four main parts to successfully control a drone using hand gestures:

- **Data Acquisition:** the database consisted of a total of 2400 raw data retrieving from five volunteers, 300 samples of each hand gestures. There are 21 points detected on each hand gesture. The MediaPipe framework is used to perform key point localization of 21 3D palm coordinates in the detected hand region.
- **Data Preprocessing:** Each of the 21 hand landmarks is composed of the coordinates x, y and z , with original values of these coordinates in the range 0 to 255. The coordinates x and y were normalized to [-1 to 1] by the image width and height respectively.
- **Model training and classification:** After the key points in each image in the database are normalized, the normalized data in entered to a convolutional neural network (CNN) to be trained for the classification of the hand gestures.
- **Generation of command signals for the drone**. The result of the classification system was used to send control signals to the DJI Ryze Tello Drone via Wi-Fi. The speed of the drone was set to 20 cm/s to be able to have a good control of the drone in an indoor environment.

RESULTS

In this project we used hand pose estimation algorithms to find key points features of hands from images. These key points were used as landmarks to classify hand gestures. Hand key point detection is the process of finding the joints on the finger as well as the finger-tips in a given image or frame in a real-time video. Experimental results demonstrated that the developed intelligent system was able to control the UAV in real time, based on the hand gestures inputs from the user. Figure 5 shows a real-time picture from the UAV's camera and how the hand key points are detected. To have a good control of the UAV in indoor environments, this experimental speed of the UAV was set to 20cm/sec.

Figure 5: Real-time picture from UAV camera

CONCLUSION

This project presented an application of computer vision used to control an UAV. The accuracy of the hand gesture classification was above 90%, so that users are able to easily control an UAV with hand gestures. However, the UAV using for this project is significantly sensitive to the ambient light due to its laser sensors for self-balancing and navigation, there is always a need to provide enough light while controlling the UAV.

This work does not stop from here, there are still many advances that can be made to this project so that the UAV has more accuracy and functions.

REFERENCES

[1]MediaPipeHands,GoogleGithub, https://google.github.io/mediapipe/solutions/hands.html
[2] Richard A. Bolt. Put-that-There: Voice and gesture at the graphics interface. SIGGRAPH Comput. Graph., 14(3):262–270, July 1980.
[3] .I. Pavlovic, R. Sharma, and T.S. Huang. Visual interpretation of hand gestures for human-computer interaction: a review. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(7):677–695, Jul 1997.
[4] A. Singandhupe, H. Manh, and D. Feil-Seifer, "Reliable Security Algorithm for Drones Using Individual Characteristics from and EEG signal", IEEE Open Access Journal, April 2018, doi: 10.1109/ACCESS.2018.2827362
[5] J. Vigliotta, J. Cipleu, A. Mikell, and R. Alba-Flores, "EMG Controlled Electric Wheelchair", Intelligent Systems Conference, Amsterdam, The Netherlands, September 2-3, 2021

This work was supported by NSF REU, Award # 1950207