Name: Abhijeet Pokhriyal

Date: 02/04/21

Project Title: Real-Time Human Action Recognition for Drone Control

Mentor and Company: Rick L. Hudson , Sr. Project Manager, School of Data Science UNC-Charlotte

Dates of Internship: Jan 20th to May 10th 2021

Project Objective:

Goal of the current project is to build an embedded real-time human action detection and classification system which would then be deployed on a Nvidia Jetson Nano chip and subsequently mounted on a drone. This will allow the pilot to secure the drone using gestures / actions that would be detected using camera signals in scenarios where drone loses controller contact with the pilot or in search and rescue missions.

Methodology:

Overall Approach

Human action recognition can be practiced primarily with three methods: Using

i. glove-based wearable devices [1],
ii. 3-dimensional locations of hand keypoints [2] and
iii. raw visual data [3]

The first method comes with the obligation of wearing an additional device with which lots of cables come even though it provides good results in terms of both accuracy and speed. The second, on the other hand, requires an extra step of hand-keypoints extraction, which bring additional time and computational cost.

Lastly, for (iii), only an image capturing sensor is required such as camera, infrared sensor or depth sensor, which are independent of the user.

In current project we will be adopting (iii) methodology and within (iii) we will be focusing on what has been described by [4] as 2D Discriminative Bottom Up approach.

Discriminative Bottom Up Approach

In the chosen approach, Waveshare Camera module [6] will be used to provide input image stream using Nvidia JetCam interface[7].

The system will first extract 2D pose of multiple people in an image and feed these pre-processed features as input to our classification model, which in turn will discriminate the video stream into one of the 4 dataset labels (see section : Dataset)

Dataset

Available to us is an in-house dataset which has 127 videos each of around ~3 seconds length. Each video is labelled with one of the four possible classes

1. Target

2. Forward
3. Descend
4. Stop

If needed, we will be leveraging MPII Human Pose Dataset [5] as well to finetune our models.

Major Tasks: (in major chunks that may be ~ 2 weeks of work and takes you to end of internship)

Jan 18- Jan 25 : Environment setup, access to dataset and the hardware. Assembling together Camera, Wifi Modules with Nvidia Jetson Nano Platform[8].

Jan 25 - Feb 8: Exploratory analysis of dataset and researching state of the art for pose/gesture recognition. Understanding the Jetson Nano specifications including processing power and memory limitations.

Feb 8 - Feb 15: Building a data pipeline to process video stream and exploring open source projects for pose detection that can be used with drone imagery. Performance test Jetson Nano with the selected open source project.

Feb 15 - March 1: Employing Machine learning to build model using Transfer learning for pose classification. Model debugging, discussing results with mentor / team.

March 1- March 15: Initial deployment of model onto Nvidia jetson to assess performance and limitations of embedded system.

March 15- March 29 : Back to drawing board to incorporate required changes. Upgrade Jetson to superior model if needed.

March 29- April 12: Finalizing Model architecture and deployment onto Jetson.

April 12- April 26: Building drone control interface with static movement commands
April 26- May 10 : Putting together the drone control interface and the machine learning model and test flight.

Please find in appendix a visual summary/Gantt chart.

Outcomes Expected:

In real-time gesture recognition applications, there are several characteristics that the system needs to satisfy:

i. An acceptable classification accuracy,
ii. fast reaction time,
iii. memory and compute efficiency

For this project the expectation is to demonstrate Computer Vision capabilities in autonomous drone control. Final product will be a fully integrated system which efficiently utilizes the embedded hardware to achieve acceptable classification accuracy in real-time. We will be demonstrating the system with a successful test flight of the drone.

Potential Risks and Strategies to Overcome :

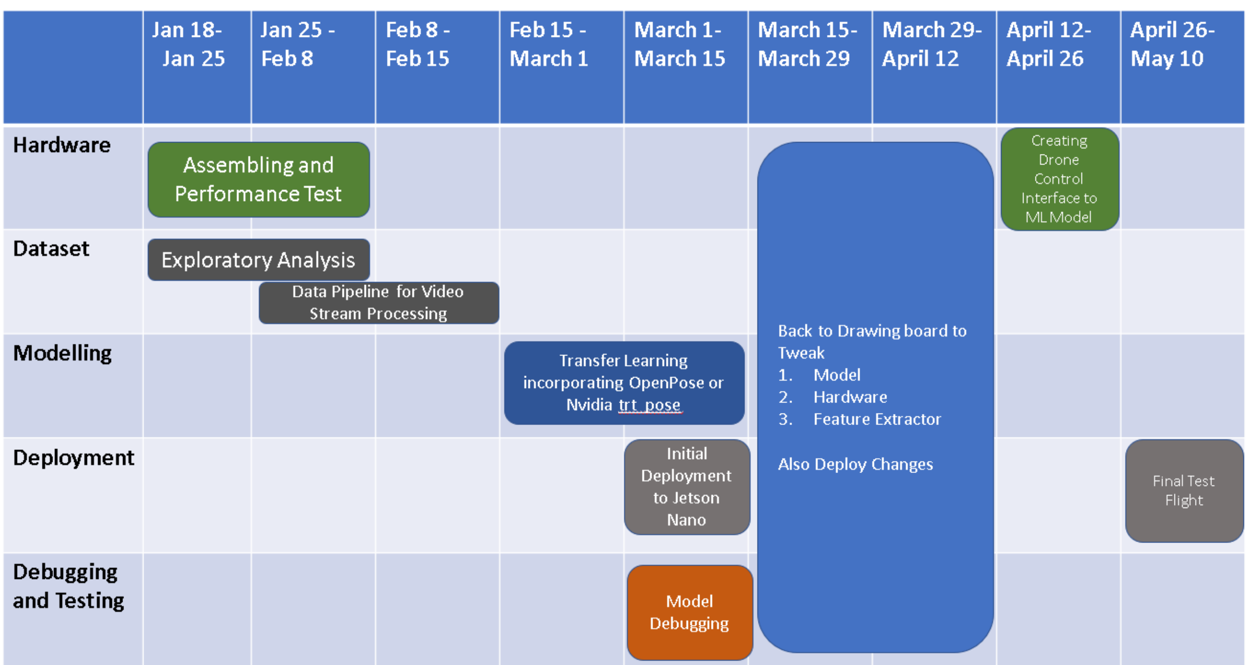
One of the major risks /bottlenecks will be the processing and memory limitations of the Nvidia Jetson Nano board. Getting to run machine learning models efficiently on an embedded system brings it's own challenges that would need to be addressed.

Another challenge would be to integrate the machine learning model with the drone control as it is something that the author will attempt for the first time.

To mitigate the aforementioned challenges, we have scheduled performance tests for the Jetson Board early on so that we can quickly identify if an upgrade is needed for the hardware. We will additionally be connecting with an another team to leverage their understanding of human computer interfaces to build our model-drone interface.

Appendix:

Gantt Chart:



SOURCES:

[1] Shin, Sungho, and Wonyong Sung. "Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks." *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2016.

[2] Han, Zhishuai, Xiaojuan Ban, and Xiaokun Wang. "Robust and customized methods for real-time hand gesture recognition under object-occlusion." *arXiv preprint arXiv:1809.05911* (2018).

[3] Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." *IEEE transactions on pattern analysis and machine intelligence* 43.1 (2019): 172-186.

[4] Chen, Y., Tian, Y., & He, M. (2020). *Monocular human pose estimation: A survey of deep learning-based methods. Computer Vision and Image Understanding, 192, 102897.* doi:10.1016/j.cviu.2019.102897

[5] http://human-pose.mpi-inf.mpg.de/

[6] https://www.waveshare.com/RPi-Camera-V2.htm

[7] https://github.com/NVIDIA-AI-IOT/jetcam

[8] https://developer.nvidia.com/EMBEDDED/jetson-nano-developer-kit