Madras Institute of Technology
Anna University, Chromepet, Chennai- 600044

DISSERTATION – II (REVIEW 2)

TARGET DETECTION AND MEASURING DIMENSIONS USING VISION TRANSFORMER FOR DRONE LANDING

PRESENTED BY:

M. VIJAYA PRAKASH (2022603026)
M.E. Avionics

UNDER THE GUIDANCE OF:

Dr. G. ANITHA
Professor

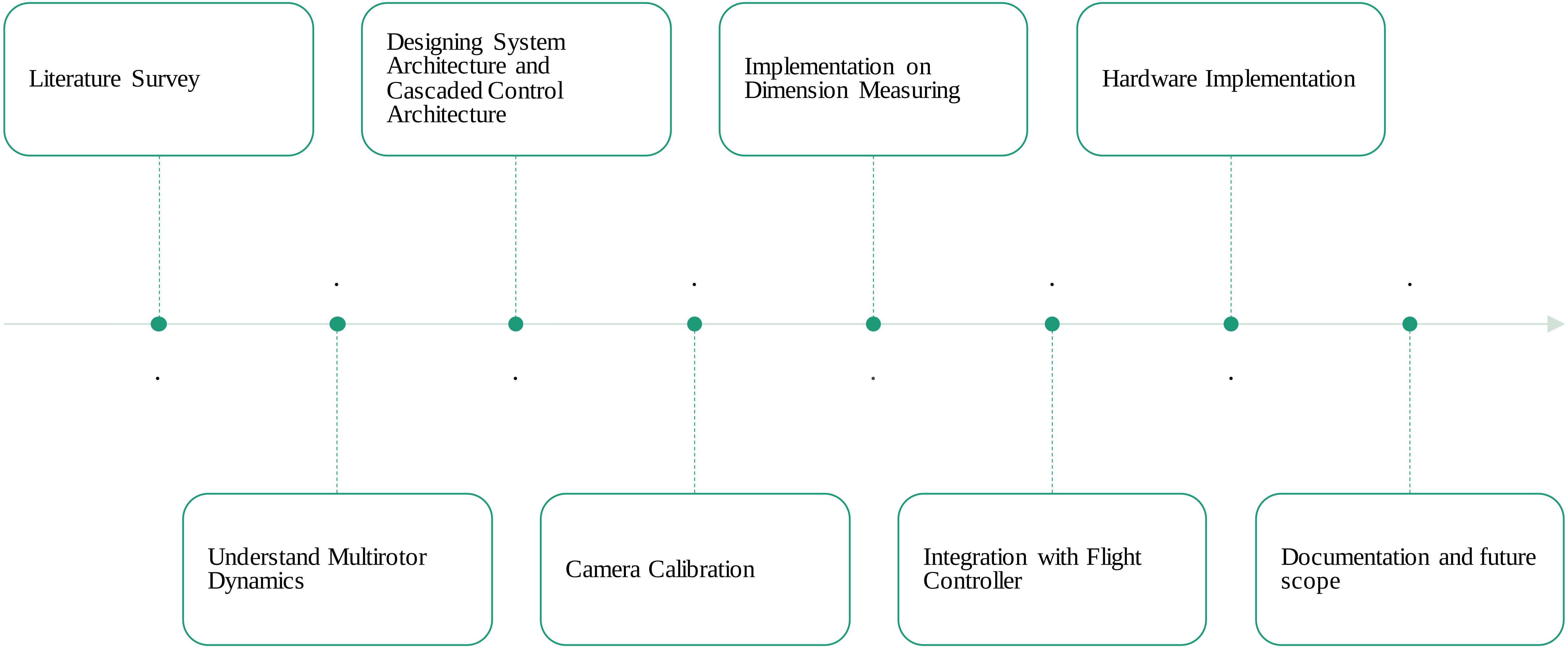
OBJECTIVE FOR DESIGNING ViT BASED DRONE LANDING

- **OBJECTIVE -** Implantation of Vision Transformer model with self-attention algorithm in drone will be performed for target object dimension measuring in real time which results in change in drone dynamics for smooth landing.
- We propose an implementation of a Vision Transformer based model, which:
  1. collects the target object data and perform preprocessing for self-attention model;
  2. is trained with relative to the pre-trained model;
  3. handles time-varying dynamic real time implementation.

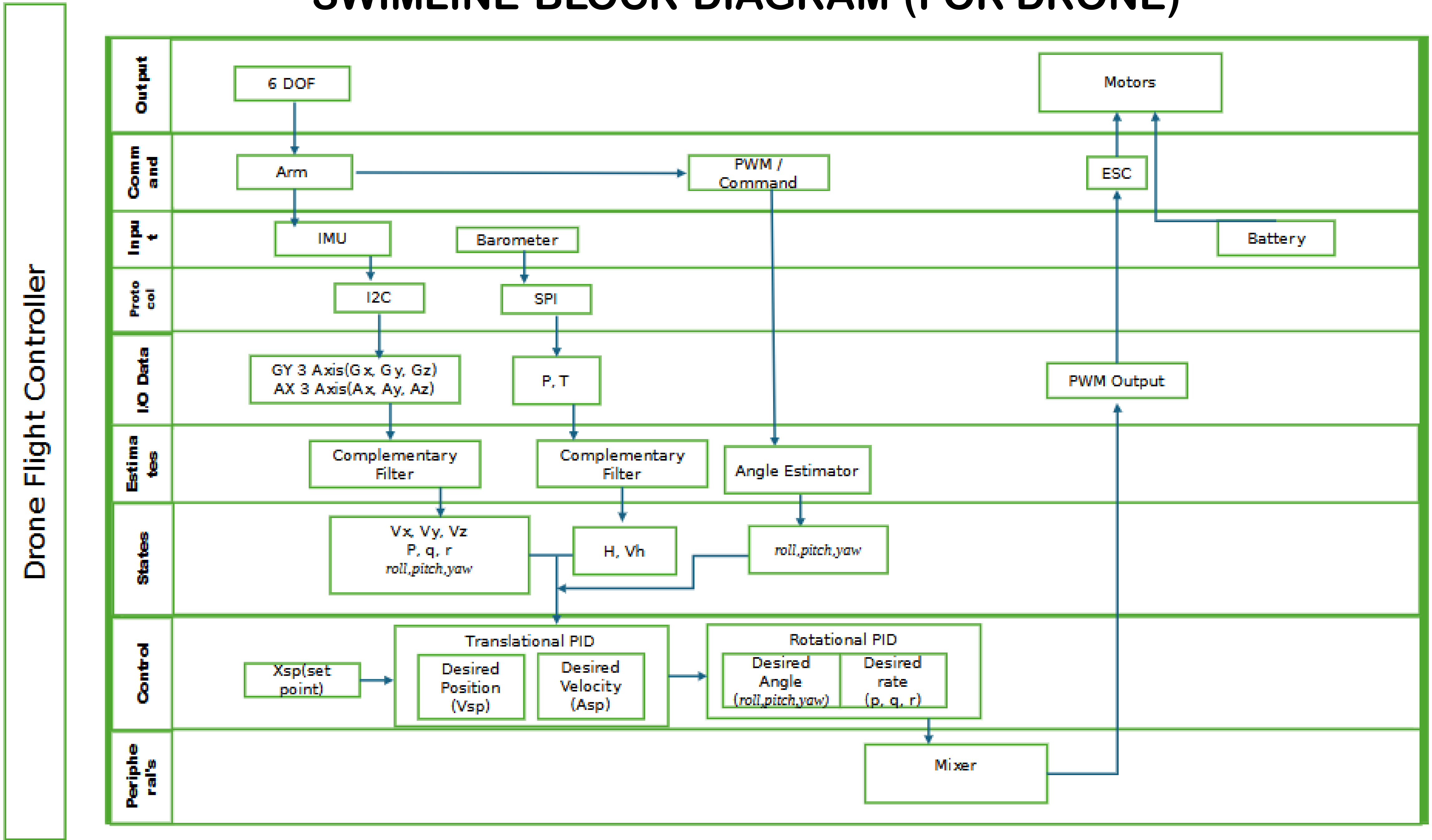
- Research Gap:
  1. Until today, Vision Transformer has often only been used to solve individual photos, but never to provide a time-varying dynamic real time image processing.
  2. Investigate the transfer learning capabilities of Vision Transformers in the process of changing drone landing in suitable environment.

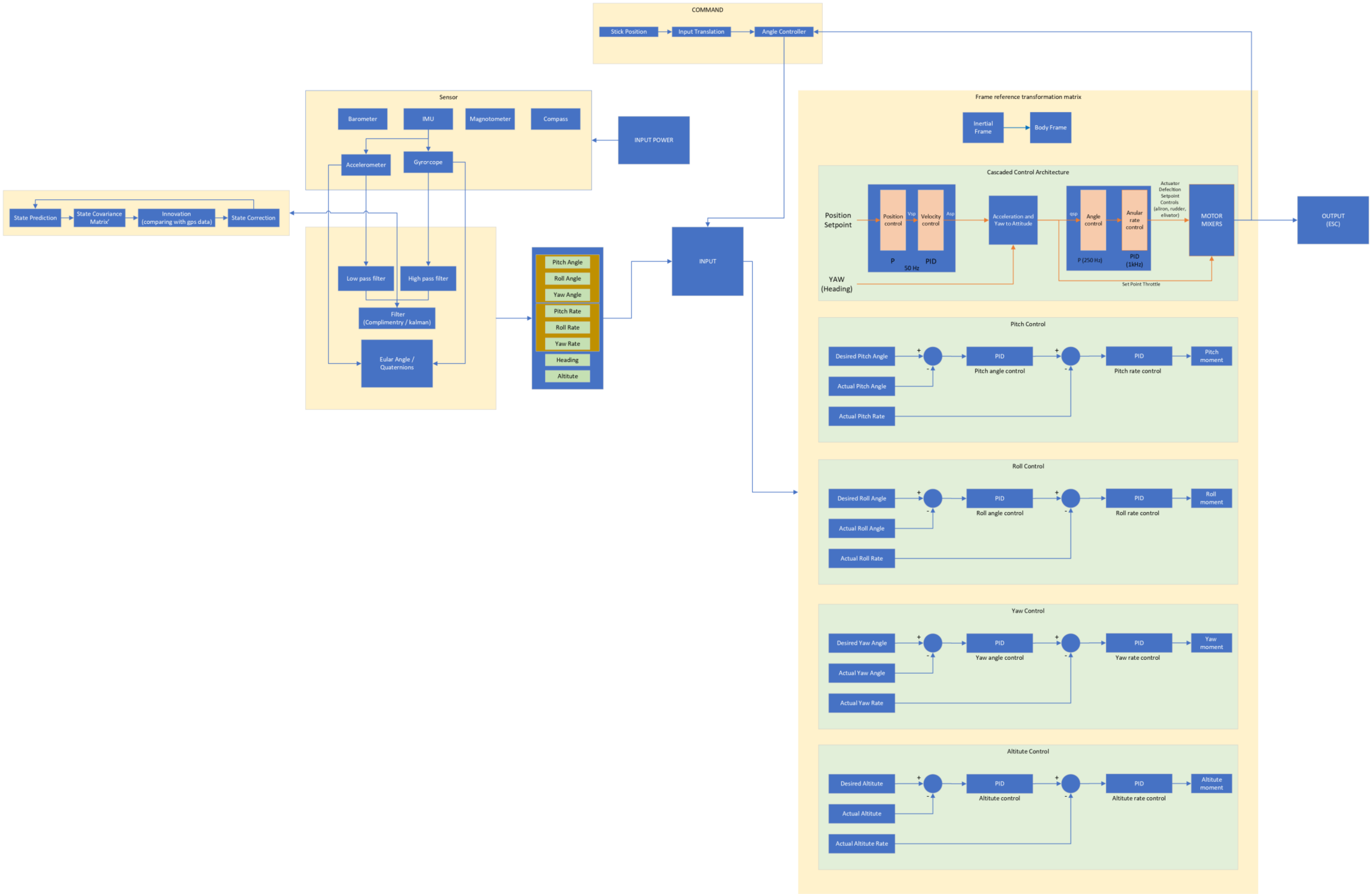
- Execution Workplan:
  1. Plan of execution will be developing ViT model for target object detection and measure its dimension. It vastly simplifies the tuning of the training, since the tuning parameters are costs that directly related to the accuracy of the training model.
  2. To achieve this, we leverage the self-attention model to implement an independent self-learning vision transformer model.
  3. The tuned data will act as an input for drone for smooth autonomous landing.

WORKFLOW


Literature Survey
Understand Multirotor Dynamics
Designing System Architecture and Cascaded Control Architecture
Camera Calibration
Implementation on Dimension Measuring
Integration with Flight Controller
Hardware Implementation
Documentation and future scope

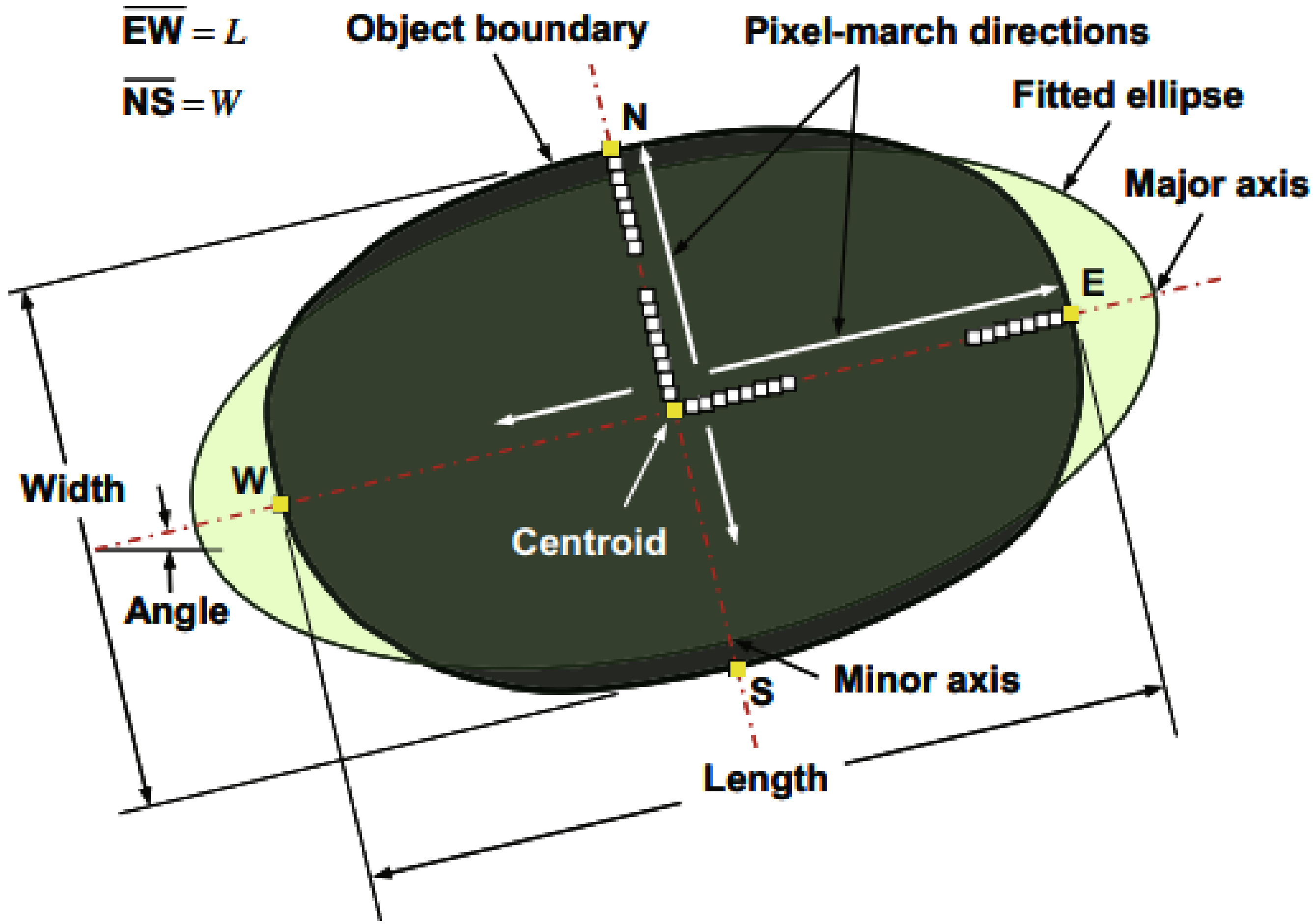
SWIMLINE BLOCK DIAGRAM (FOR DRONE)

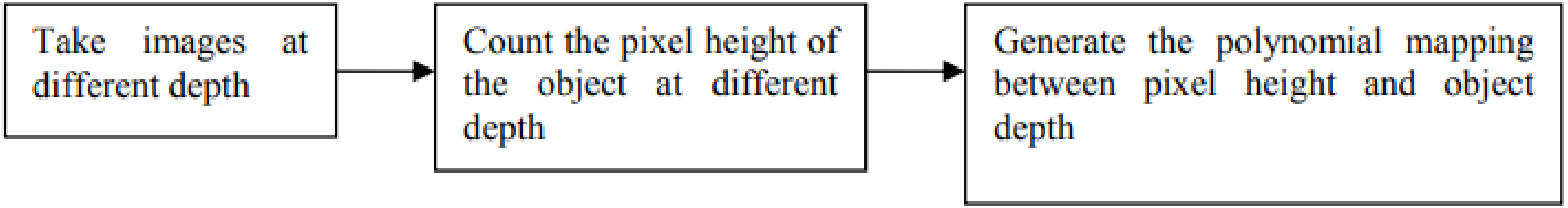




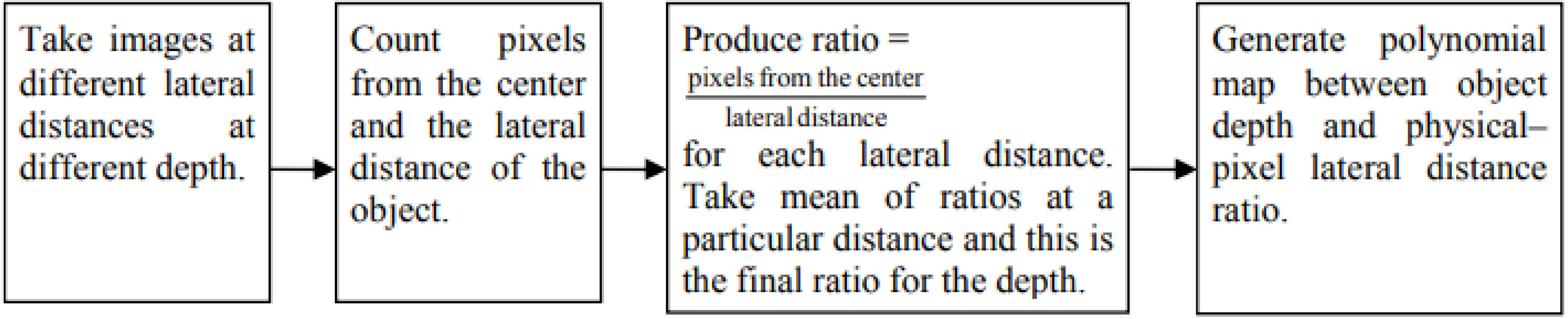
COMMAND
Stick Position
Input Translation
Angle Controller
Sensor
Barometer
IMU
Magnotometer
Compass
INPUT POWER
Accelerometer
Gyroscope
State Prediction
State Covariance Matrix'
Innovation (comparing with gps data)
State Correction
Low pass filter
High pass filter
Filter (Complimentry / kalman)
Eular Angle / Quaternions
Pitch Angle
Roll Angle
Yaw Angle
Pitch Rate
Roll Rate
Yaw Rate
Heading
Altitute
INPUT
Frame reference transformation matrix
Inertial Frame
Body Frame
Cascaded Control Architecture
Position Setpoint
Position control
Vsp
Velocity control
Asp
P
PID
50 Hz
Acceleration and Yaw to Attitude
qsp
Angle control
Anular rate control
P (250 Hz)
PID (1kHz)
Actuator Defecltion Setpoint Controls (aliron, rudder, elivator)
MOTOR MIXERS
YAW (Heading)
Set Point Throttle
OUTPUT (ESC)
Pitch Control
Desired Pitch Angle
Actual Pitch Angle
Actual Pitch Rate
PID
Pitch angle control
PID
Pitch rate control
Pitch moment
Roll Control
Desired Roll Angle
Actual Roll Angle
Actual Roll Rate
PID
Roll angle control
PID
Roll rate control
Roll moment
Yaw Control
Desired Yaw Angle
Actual Yaw Angle
Actual Yaw Rate
PID
Yaw angle control
PID
Yaw rate control
Yaw moment
Altitute Control
Desired Altitute
Actual Altitute
Actual Altitute Rate
PID
Altitute control
PID
Altitute rate control
Altitute moment

PIXEL-MARCH METHOD.



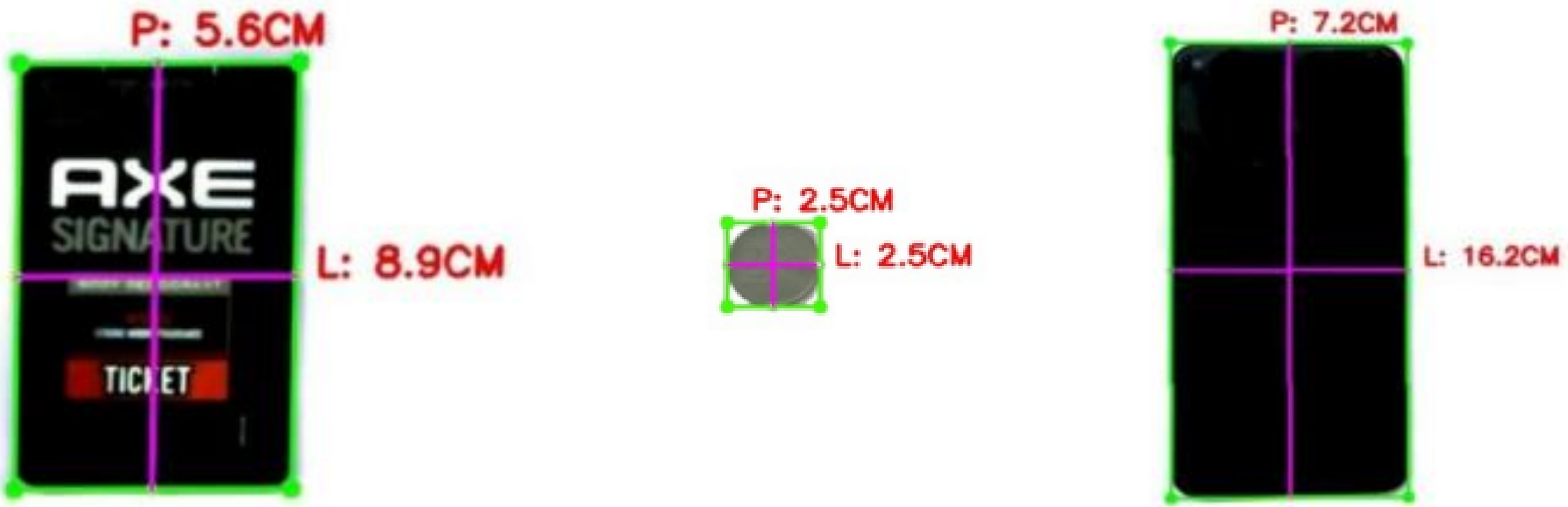


(a) Learning a map between object depth and pixel height

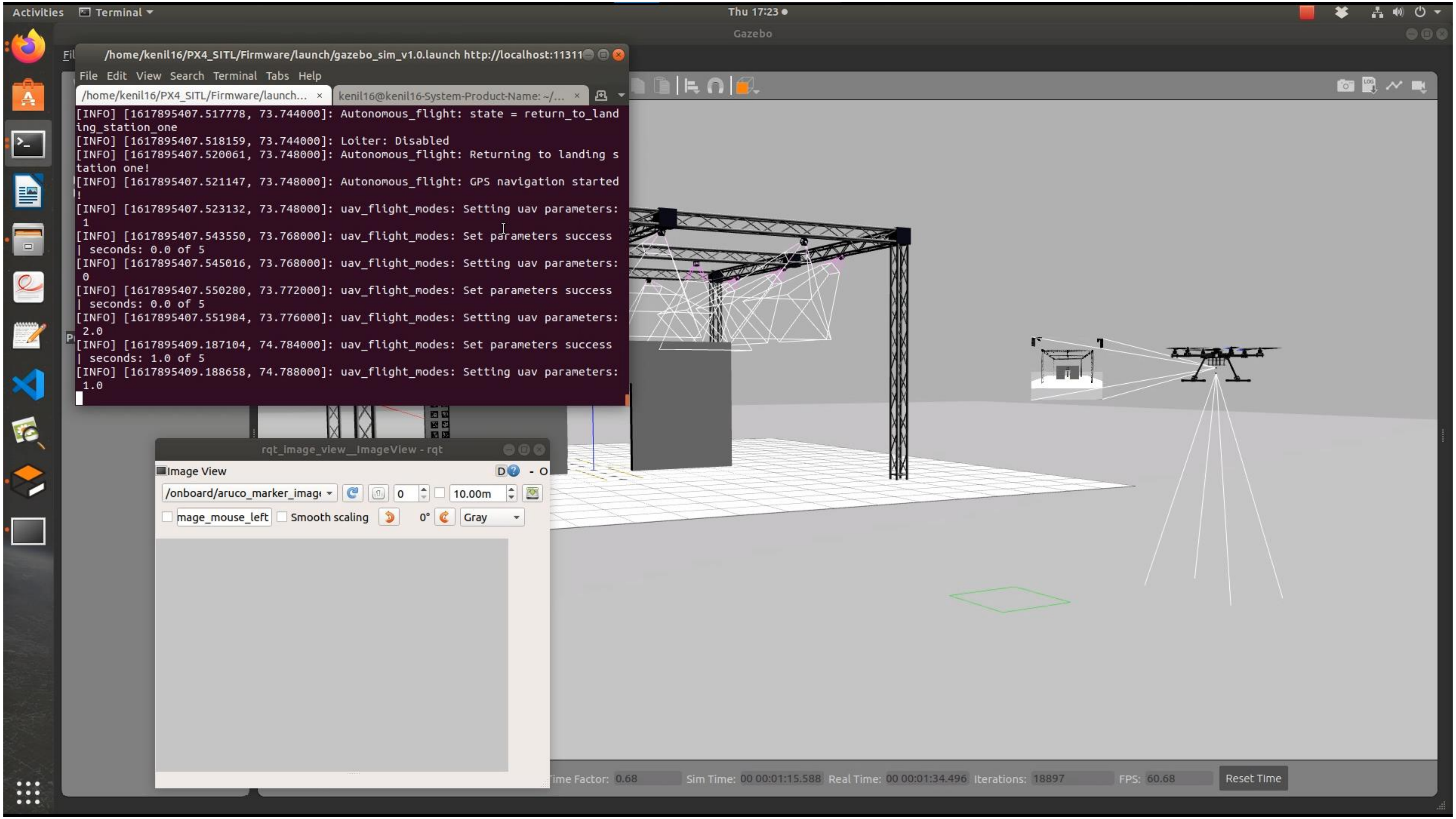


(b) Learning a map between object depth and physical–pixel lateral distance ratio

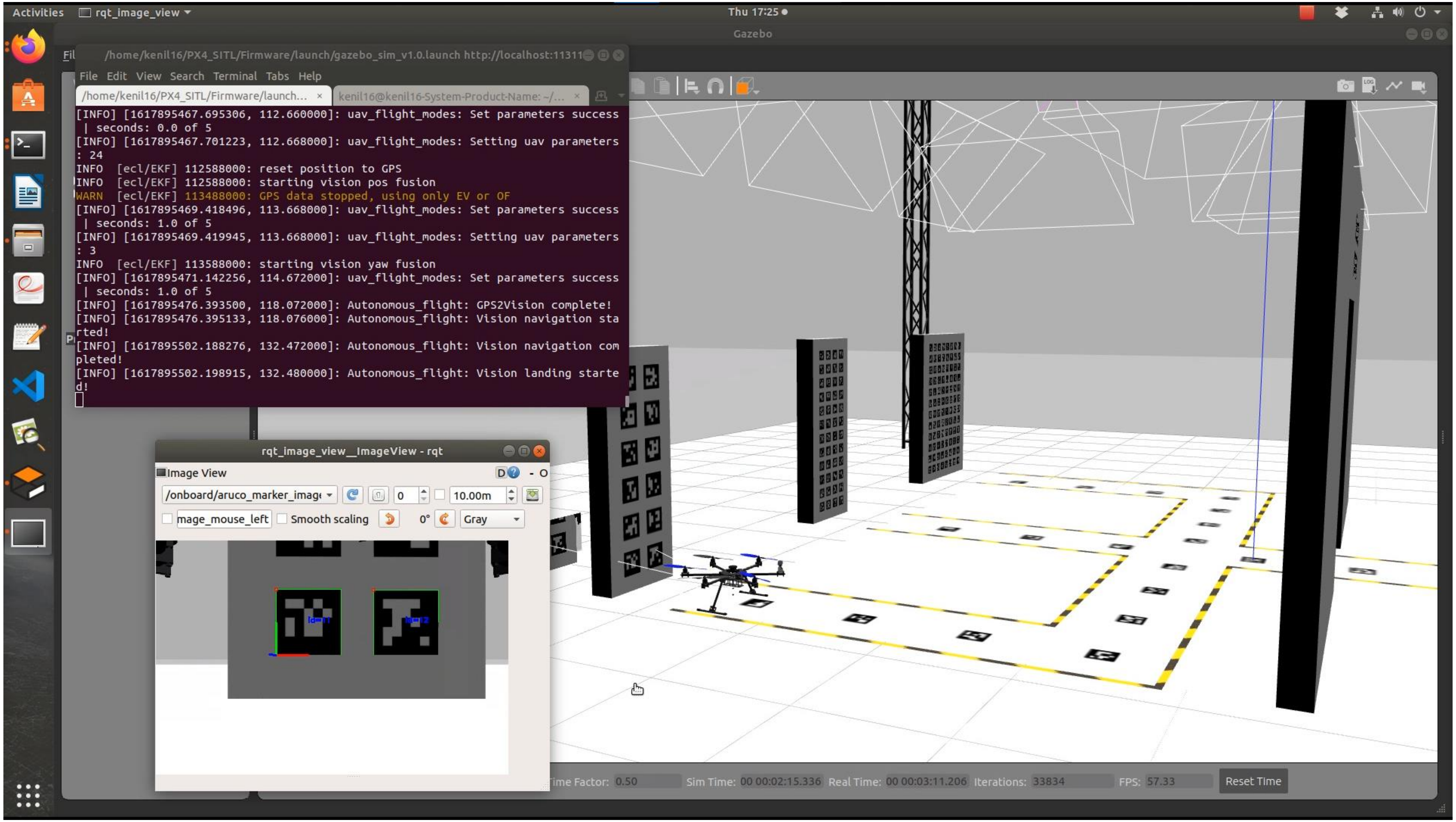
OUTPUT - DIMENSION DETECTION (CANNY EDGE DETECTION ALGORITHM)



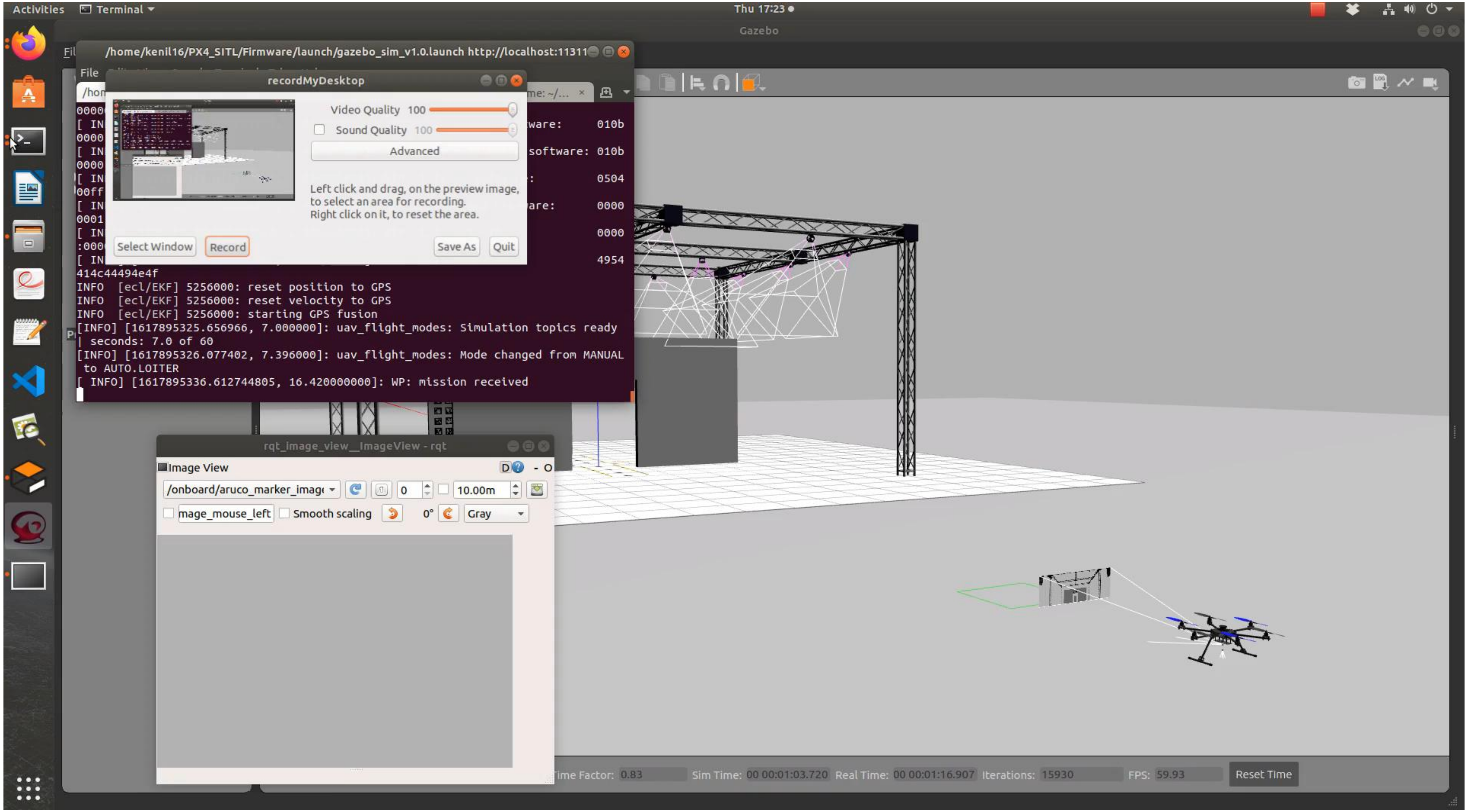
OUTPUT – TAKEOFF COMMAND



OUTPUT – LAND COMMAND

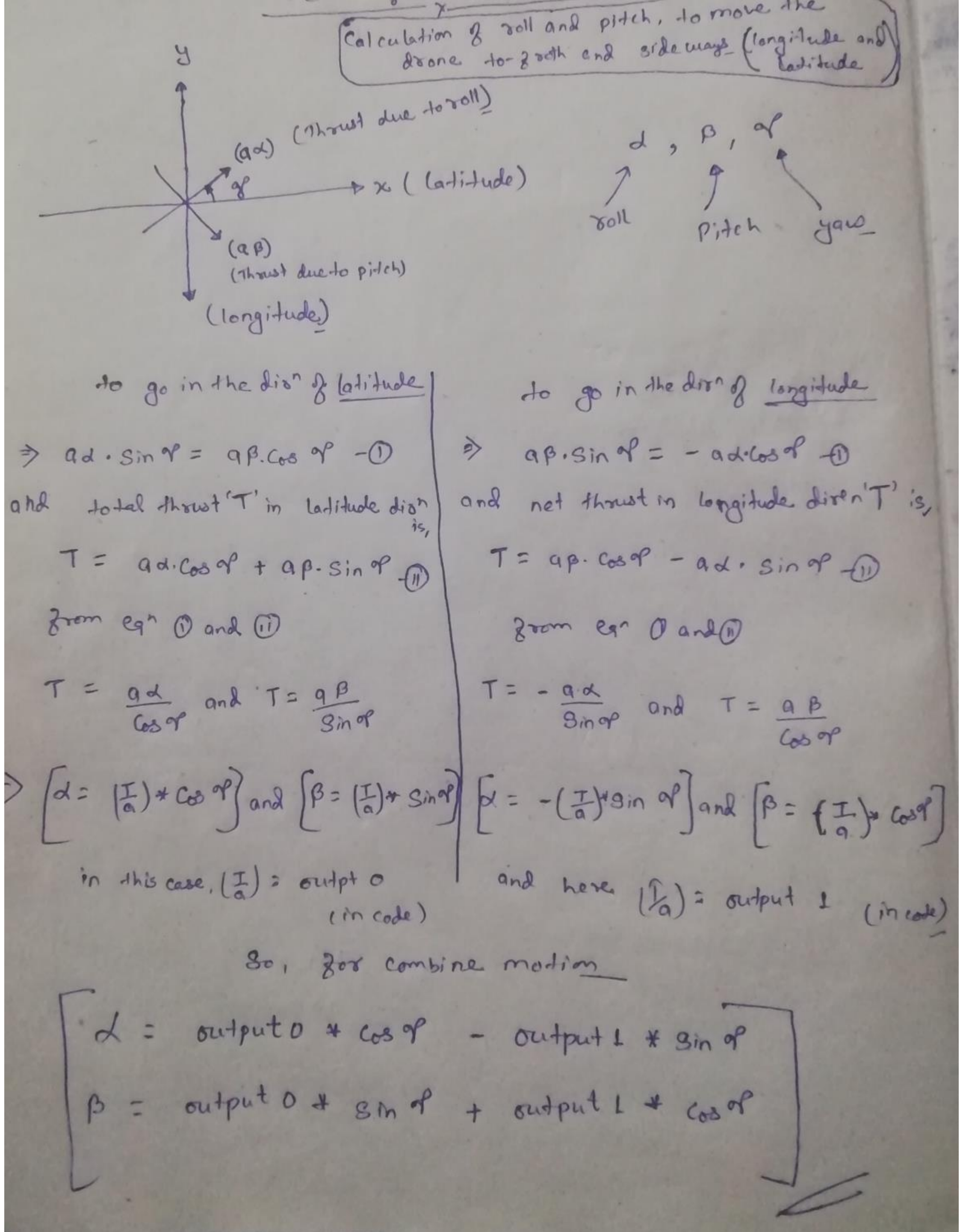


OUTPUT - NAVIGATION



OUTPUT – HARDWARE TESTING

CALCULATION OF ROLL AND PITCH TO MOVE THE DRONE IN A HORIZONTAL PLANE



Calculation of roll and pitch, to move the drone to-forth and sideways (longitude and latitude)

to go in the dirⁿ of latitude

$\Rightarrow a\alpha \cdot \sin\gamma = a\beta \cdot \cos\gamma$ —(i)

and total thrust 'T' in latitude dirⁿ is,

$T = a\alpha \cdot \cos\gamma + a\beta \cdot \sin\gamma$ —(ii)

from eqⁿ (i) and (ii)

$T = \frac{a\alpha}{\cos\gamma}$ and $T = \frac{a\beta}{\sin\gamma}$

$\Rightarrow \left[\alpha = \left(\frac{T}{a}\right) * \cos\gamma\right]$ and $\left[\beta = \left(\frac{T}{a}\right) * \sin\gamma\right]$

in this case, $\left(\frac{T}{a}\right)$ = output 0 (in code)

to go in the dirⁿ of longitude

$\Rightarrow a\beta \cdot \sin\gamma = -a\alpha \cdot \cos\gamma$ —(i)

and net thrust in longitude dirⁿ 'T' is,

$T = a\beta \cdot \cos\gamma - a\alpha \cdot \sin\gamma$ —(ii)

from eqⁿ (i) and (ii)

$T = -\frac{a\alpha}{\sin\gamma}$ and $T = \frac{a\beta}{\cos\gamma}$

$\left[\alpha = -\left(\frac{T}{a}\right) * \sin\gamma\right]$ and $\left[\beta = \left(\frac{T}{a}\right) * \cos\gamma\right]$

and here $\left(\frac{T}{a}\right)$ = output 1 (in code)

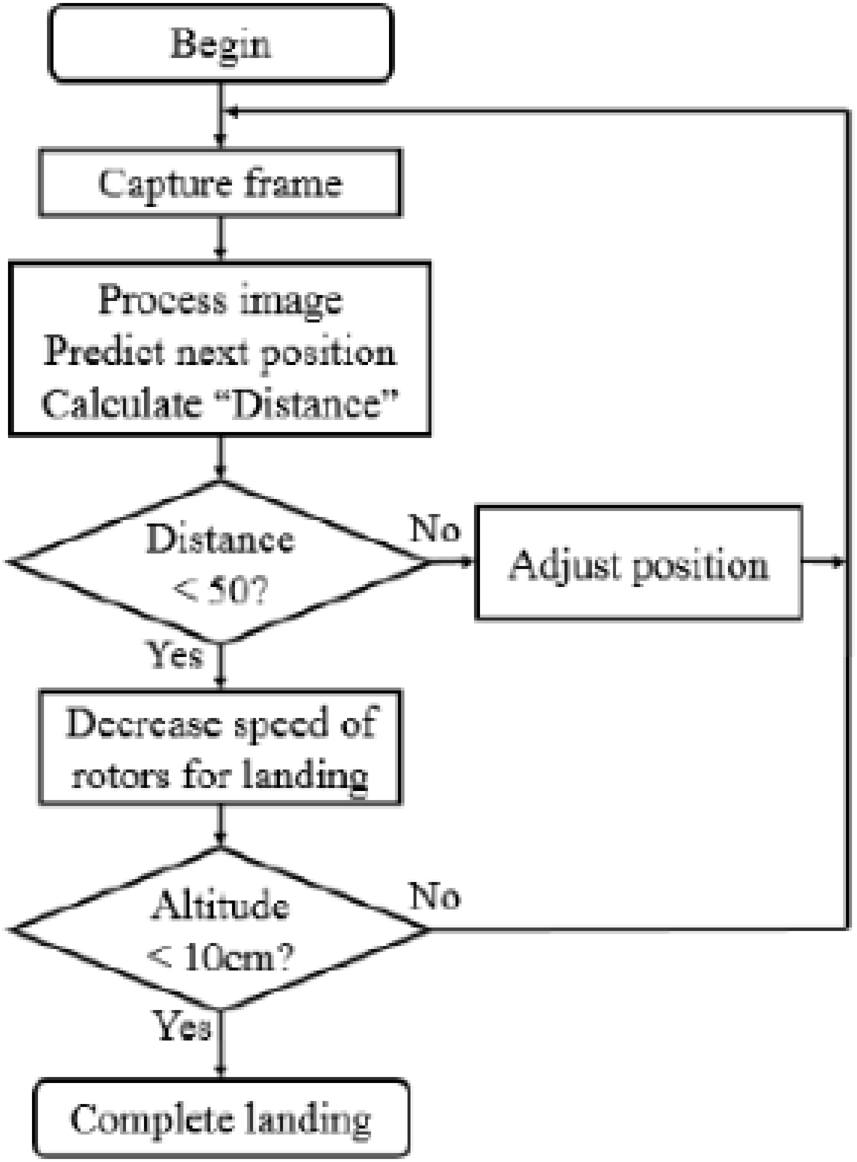
So, for combine motion

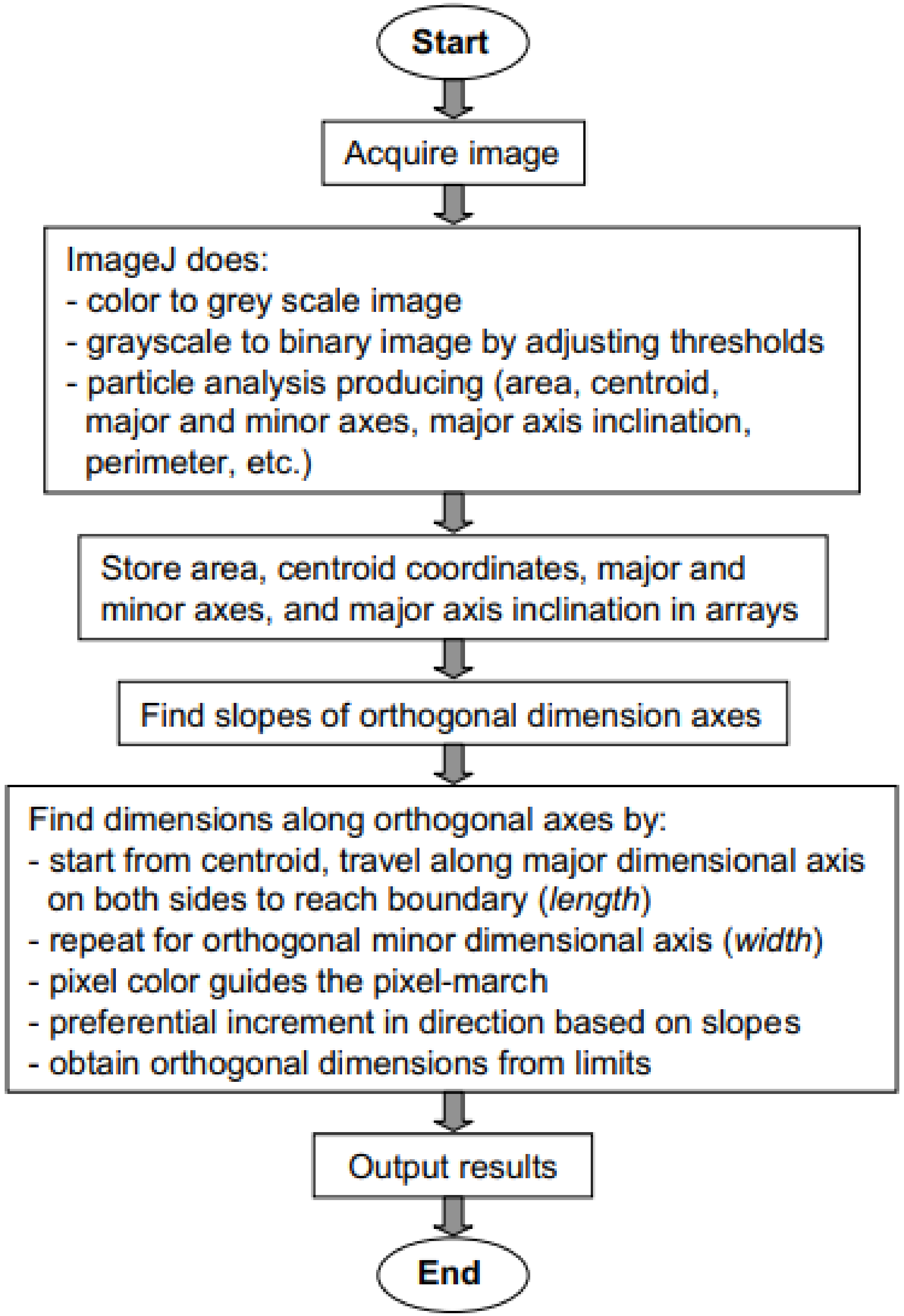
$$\begin{bmatrix} \alpha = \text{output0} * \cos\gamma - \text{output1} * \sin\gamma \\ \beta = \text{output0} * \sin\gamma + \text{output1} * \cos\gamma \end{bmatrix}$$

Objective -

1. Pitch calculation
2. Roll calculation
3. To move aerial vehicle from one point to another point in horizontal plane

METHODOLOGY





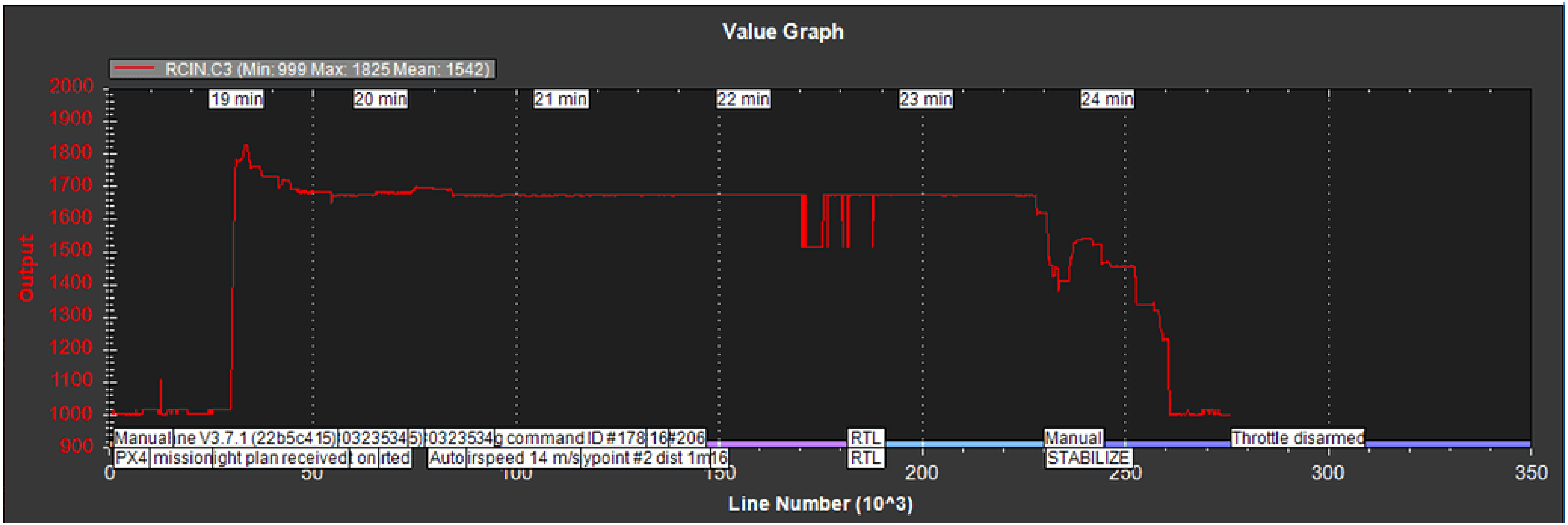
OUTPUT – NAVIGATION THROUGH COMPUTER VISION

AUTOMOUS LANDING THROUGH CAMERA INPUT (4 STAGES)

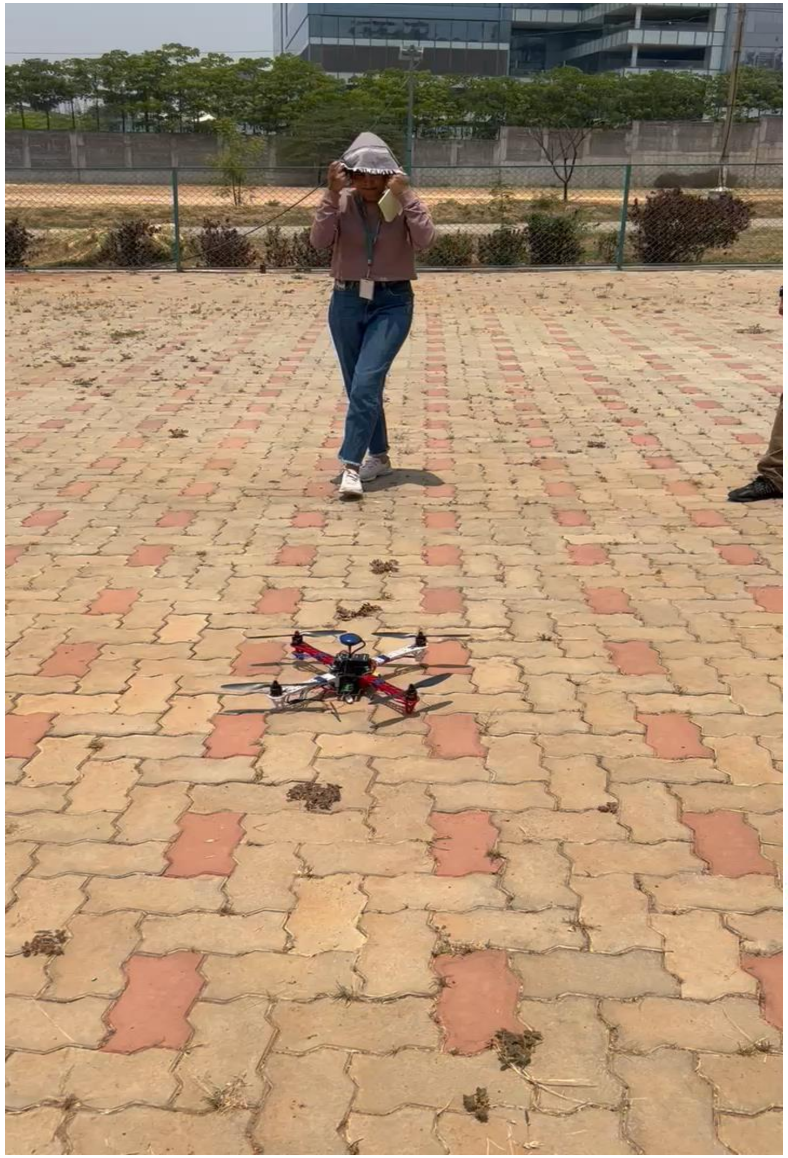
TAKE-OFF -> CRUISE -> LANDING DETECTION (THROUGH CAMERA) -> LANDING TRIGGERED

OUTPUT (ACTUAL AND PREDICTED LANDING)

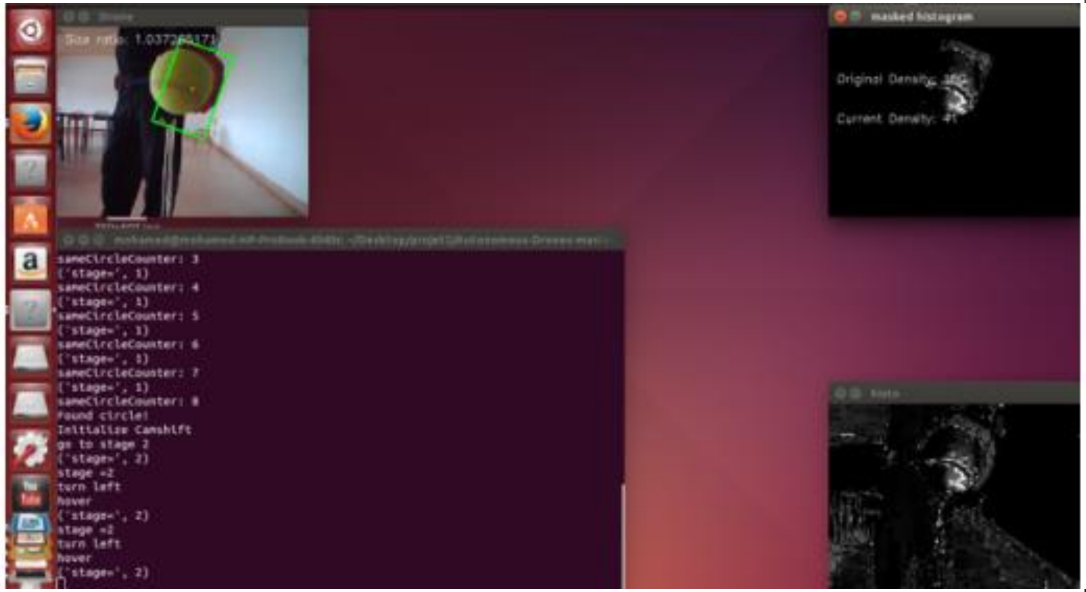
RC LOG DATA THRUGHT THE OPERATION



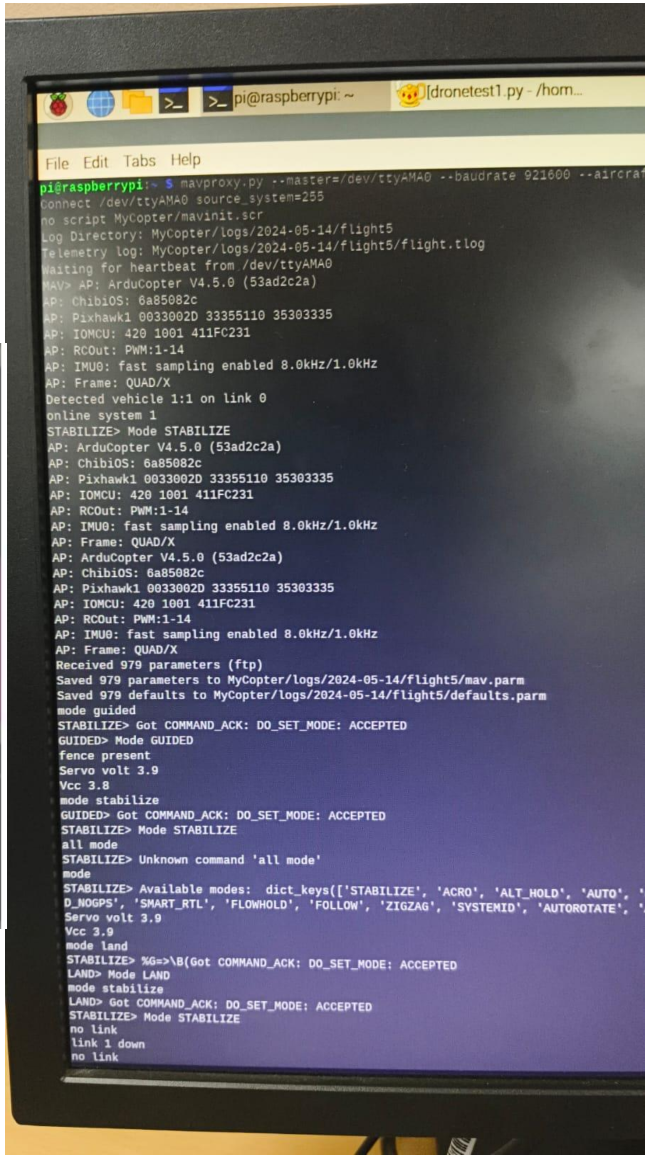
ONFIELD TESTING



ONFIELD TESTING – WITHOUT CV (JUST LANDING COMMAND)



COLOUR GRADIENT TESTING



PIXHAWK MAVLINK CONNECTION TESTING

FUTURE WORKS

- It will be implemented with small category (weight) drone.
- This tech will be integrated with the distance sensor module for short distance measurement.
- Improvement in colour gradient landing will be done.

REFERENCE

1. Lecture #4: Pixels and Filters, Brian Hicks, Alec Arshavsky, Sam Trautwein, Christine Phan, James Ortiz, Department of Computer Science, Stanford University

2. Tutorial 6: Transformers and Multi-Head Attention — UvA DL Notebooks v1.2 documentation (uvadlc-notebooks.readthedocs.io)

3. How to Train the Hugging Face Vision Transformer On a Custom Dataset (roboflow.com)

4. Visual Saliency Transformer

5. Salient Object Detection: A Discriminative Regional Feature Integration Approach

6. Indoor versus outdoor scene recognition for navigation of a micro aerial vehicle using spatial color gist wavelet descriptors

7. Singh, Ankur & Gupta, Anurag & Gupta, Amit & Chaudhary, Archit & Jhamb, Bhuvan & Sahil, Mohd & Saraswati, Samir. (2023). Architecture and Algorithms for a Pixhawk-Based Autonomous Vehicle. 10.1007/978-981-99-0236-1_34.

8. Meier, Lorenz & Tanskanen, Petri & Heng, Lionel & Lee, Gim & Fraundorfer, Friedrich & Pollefeys, Marc. (2012). PIXHAWK: A micro aerial vehicle design for autonomous flight using onboard computer vision. Autonomous Robots. 33. 10.1007/s10514-012-9281-4.

THANK YOU