Augmented Aerial Swarm Behavior via Natural Human Interaction

Amogh Bhagwat[1], Akshay T[1], Atharva Jaipurkar[1], Ayush Khandelwal[1], Devang Sailor[1], Harshal Kataria[1], Kaushal Kishore[1], Prasanth YSS[1], Prashant K. Sharma[1], Shinjan Mitra[1], Trishant Roy[1], Vikrant Nagpure[1], and Vivek Sangwan[2]

[1]Members, Unmesh Mashruwala Innovation Cell, IIT Bombay
[2]Assistant Professor, Department of Mechanical Engineering, IIT Bombay

ABSTRACT

The International Aerial Robotics Competition, in its 8$^{\text{th}}$ Mission statement demands the completion of a task by a human player assisted by a swarm of 4 aerial vehicles while avoiding laser hits from 4 enemy aerial sentries. An interlinked system is designed between the aerial vehicles, an off-board computational unit and the player equipped with a portable device. The robots are made fully aware of their surroundings via human and object detection, localization, and target identification. This paper presents a detailed description of the various modules developed, and how they fit together to enable swarm behavior.

INTRODUCTION

State-of-the art autonomous aerial systems are becoming increasingly commonplace due to the advances in positioning and control systems. Coupled with the strides in the field of machine learning and artificial intelligence, aerial robots can now be commercially purchased that can follow a human while recording live video. In the quest to further push the limits of flying machines, the International Aerial Robotics Competition (IARC) comes up with unique and seemingly *impossible* challenges. So far teams have already demonstrated autonomous flight, indoor navigation in GPS-denied environments, obstacle identification and game-based challenge completion.

In the 8$^{\text{th}}$ mission of the IARC, the problem of man-machine interaction has been posed. This paper illustrates the work done by the members of IIT Bombay in pursuit of a solution to this complex task. This section elaborates on the problem statement, our approach towards the solution and the team's yearly milestones. The rest of the paper is structured as follows:

- Air Vehicle illustrates the structure and properties of the aerial vehicle including the choice of components after taking weight and thrust into account.
- In Payload, the various sensors and communication modules installed on the system are described along with the power management system.
- Flight preparation checklist and man/machine interfaces are described in Operations.

- The Risk Reduction section describes the various measures taken to ensure safety.
- Finally the findings are summarized in Conclusion

Statement of the problem
Mission 8 of the International Aerial Robotics Competition takes the challenge of building autonomous flying machines to another level. This is the first time that man-machine interaction has been introduced into the problem statement. On top of all the behaviours expected in Mission 7, the aerial vehicles must be able to respond to non-electronic human commands. The problem statement requires a human participant in the arena hereby referred to as the *player*, assisted by upto four aerial vehicles. The aim of the *player* is to retrieve a technical component from 4 locked bins. The unlock code for each bin is displayed on its top and has to be *seen* simultaneously by the aerial vehicles and transmitted to the player's hand-held device. Enemy aerial vehicles in the arena will attempt to hit the player with laser beams. The player must complete the task in under 8 minutes without getting hit by the laser 10 times. The player's *friendly* aerial assistants may shield or heal the player, while making sure they do not collide with each other or the enemy vehicles.

Conceptual solution
The approach to solve the problem involved splitting the problem into various sub-systems and tackling each individually. Once a reasonable working model of each sub-sysem is obtained, they are integrated into the system. The sub-problems consist of:

- **Electromechanics** - Construction of the aerial vehicle and power system. Also in charge of safety systems like propeller guards and kill switch.
- **Controls** - Overall flight controls which includes providing pose values to a package resulting in vehicle movement.
- **Localization** - Obtaining the current position of the vehicle at all times in a GPS-denied environment.
- **Obstacle detection and avoidance** - Identify enemy quad-copters, static objects and fellow helper robots and take evasive maneuvers if too close to them.
- **Human detection and tracking** - Identifying the human player in the arena and positioning the aerial vehicle at a fixed distance from the player.
- **Human command recognition** - Divided into audio and gesture recognition.
- **Target identification** - Locating the bins within the arena and guiding the vehicle towards it.

Yearly Milestones
As this is IIT Bombay's first participation in the IARC, the goals for this year include the development of all hardware and software elements on each aerial vehicle as per Mission 8 requirements. The target is to have four fully functional aerial vehicles with the complete suite of man-machine interactions as mandated by the problem statement.

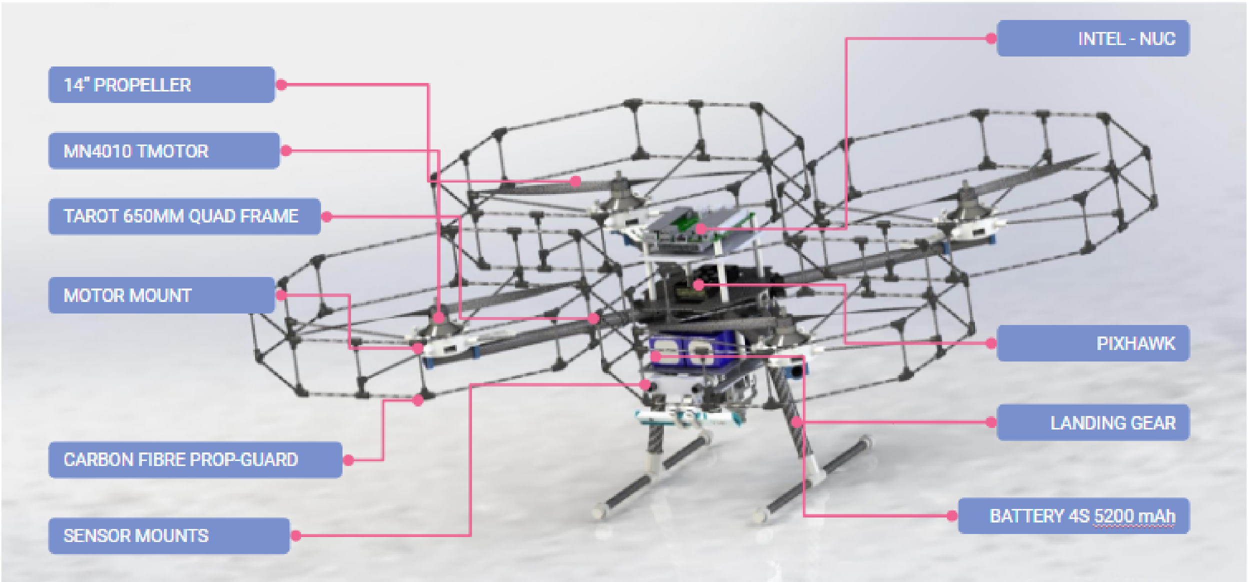

Figure 1. 3D rendered model of the aerial vehicle with major components labeled

AIR VEHICLE

A quad-copter design forms the basis of the aerial platform used to carry all the sensors and computation units as payload. The payload of 1 kg includes one NUC computational unit, D415 depth camera, LiDAR sensor, PX4Flow and 4 monocular cameras. The whole system sits on a 650 mm Tarot Iron Man X-frame. The total weight of each quad-copter is about 3 kg.
The main component of the software subsystem of the quad-copter is the decision making and controls node which integrates all the other sub-systems namely human detection and tracking, audio recognition, path planning, gesture recognition via decision making algorithms. The safety system is responsible for the immediate emergency response based on predefined conditions.

Propulsion and Lift System

As shown in Fig.1 an X frame type quad-copter has been developed. The design of this X frame provides both agility and balanced mobility. To make the system airborne, 4 high performance MN4010 tiger motors along with 14 inch propellers are used. They provide a maximum thrust of 6 kg. These motors are used to provide lift along with pitch (longitude) and roll (lateral) control.
Carbon fiber propellers have been selected over plastic variants as they don't flex and loose energy when in motion. To achieve a flight time of 10 min, two 4S 5200 mAh batteries are used to power the quad-copter and on-board sensors.

Guidance, Navigation and Control

The guidance, navigation and control (GNC) functionality is provided by the stability augmentation and navigation system. Fig.2 provides an overview of the GNC system.

Stability Augmentation System
Stability augmentation is performed by the on-board Pixhawk flight controller. The Inertial Mass Unit (IMU), LIDAR Lite v3 and PX4Flow allow it to respond to vehicle instabilities and maintain a stable flight. The use of extended Kalman filters on the sensors data and PID controllers add further stability. The quad-copter's stability has been attained through repeated testing, calibration, and machine learning.
The MAVROS ROS package enables MAVLink-extendable communication between computers running ROS, MAVLink enabled autopilots, and MAVLink enabled GCS.
Mechanically, the quad-copter's X configuration provides flexibility in control. The weight distribution of the quad-copter is such that its center of mass lies at its geometric center in the x-y plane and below the propellers in the z-axis. The Risk Reduction section discusses the vibration reduction measures which further aid in achieving the aerodynamic stability of the quad-copter.

Navigation
The system comprises of localization and obstacle avoidance

- For indoor localization of the quad-copter,optical flow and visual odometry methods have been used. They are based on PX4Flow optical flow sensor and front facing stereo camera respectively. The sensor data is fed into an EKF/UKF state estimator that gives the estimated quad-copter pose.

  The environment is sensed via a front facing depth camera and feature extraction and matching is used to get reliable pose value. PX4Flow optical flow sensor has been used which is interfaced to the on-board computer through a USB.

  The PX4Flow node is used to get the optical values that are fed into the state estimation package. The quad-copter's position given by optical flow is further corrected by that obtained using visual odometry methods.

- Quad-copter navigation is achieved by assuming a straight line path between the initial and final positions. Any obstacles encountered in the path are avoided using an artificial potential field-based approach. It is a real-time robot path planning method, and is widely used for autonomous mobile robot obstacle avoidance due to its elegant mathematical analysis and simplicity. The approach considers the goal as a generator of attractive potential gradient while the obstacles generate repulsive potential gradient in a predefined radial vicinity d^*. The attractive and repulsive potential gradients are given by

  $$\nabla U_{att} = \begin{cases} \alpha(q - q_{goal}) & if\ \|q, q_{goal}\| < d \\ d\alpha(q - q_{goal}) / \|q, q_{goal}\| & if\ \|q, q_{goal}\| > d \end{cases}$$

  $$\nabla U_{rep} = \beta(1/d^* - 1/d_{obs}) * 1/d_{obs}^2 * \nabla d_{obs}$$

  where,
  q is the position of the quad-copter, q_{goal} is position of goal , α and β are the scaling factors and d is the quadratic threshold for attractive potential.
  Total obstacle and goal effect is sum of gradient of individual obstacle and goal. The quad-copter moves in the direction opposite of maximum gradient to reach the goal.

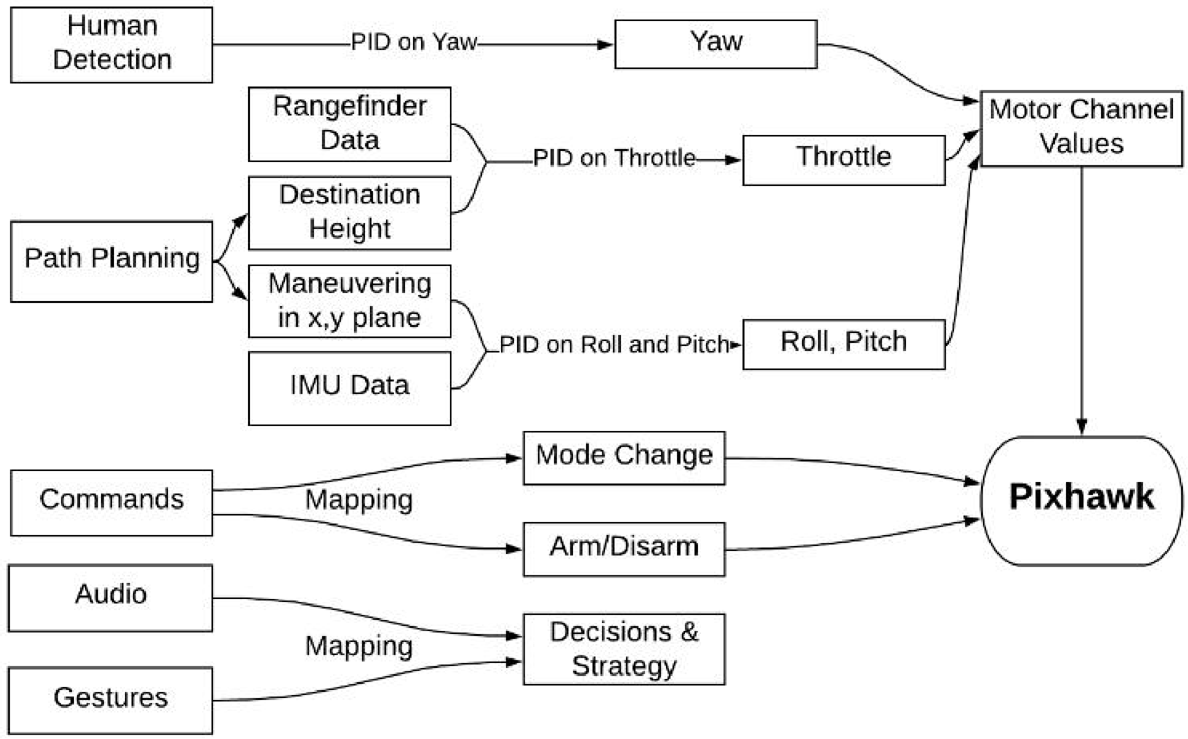

Figure 2. Control System Architecture

Flight Termination System

The flight termination system is a key component of the Stability Augmentation System. It allows the quad-copter to take predefined actions, while providing the option for a human user to take control, if needed. A flight planning ROS node runs recurring analysis and diagnostics using the data stored in the software and hardware databases. It initiates controlled landing in following cases:

- It detects anomalies based on a set of preloaded anomaly signatures
- 8 minutes have elapsed since takeoff
- Battery voltage of the quad-copter drops below 14.9V
- Command executed through ground station

Even if the software flight termination stack fails, power to the motors can be cut via the kill switch as explained in Safety causing the vehicle to shut down and land. This kill switch can be activated from the base station in the event of an anomaly. Given the capabilities of the landing gear, from operating heights of upto 2m this will not cause any harm to the quad-copter or its surrounding environment.

PAYLOAD

Sensor Suite

Guidance Navigation and Control

Pixhawk flight controller has two on-board Inertial Measurement Units, one 3 axis and another 6 axis. It takes feedback from these sensors to maintain its 3D orientation. A LiDAR-Lite v3 altimeter that works up to a maximum altitude of 40m has been used and it helps the quad-copter hold the specified altitude. An Optical Flow camera has also been employed. It has a native resolution of 752480 pixels and calculates optical flow on a 4x binned and cropped area at 400 Hz, making it highly sensitive to light. It enables the quad-copter to hold its position in the Cartesian coordinates by detecting change in the features of the image. It works indoors and also in low light outdoor conditions.

Mission Sensors

The main mission-critical sensors are:

- Intel D415 Stereo Camera for front view used for navigation and object detection
- Downward facing Monocular Camera for video feed of unlock code
- Backward facing Monocular camera for human tracking
- 2 sideways facing monocular camera for 3D obstacle detection and avoidance

All the cameras are connected to the on-board computer and their video feeds are sent to ground station as well as to the player and judges. At ground station the video feeds are processed using custom software, machine learning approaches and OpenCV framework.

Target Identification

For the identification of our targets i.e. the boxes in the arena, letters on these boxes (i.e "I", "A", "R", "C") and enemy quad-copters, a CNN-based architecture [1] has been used. Generally, the detection systems use the approach of applying the model at multiple location of the image and taking the high scoring regions into consideration. A much better way is to apply the neural network to the full image at once. In this method, the image is divided into regions and prediction for the bounding boxes as well as probability of the regions is obtained. The bounding boxes are weighted by the estimated probabilities.
This approach has several advantages over the conventional classifier-based systems. The network considers the whole image at the test time. Thus predictions made consider the global context in the image.
Unlike the network architectures like [2], which require thousands of network evaluation to make good predictions, this architecture needs only one network evaluation. Thus, making it almost 1000 times faster than [2] and approximately 100 times faster than [3].

Approximately 2000 samples have been used for each: boxes and quad-copters for training, validation and testing. The dataset is prepared by the team, with varying distances, angles and height. Also, video feed from the quad-copter is used for the dataset. This allowed to capture various scenarios, allowing it to work robustly in real world scenarios.

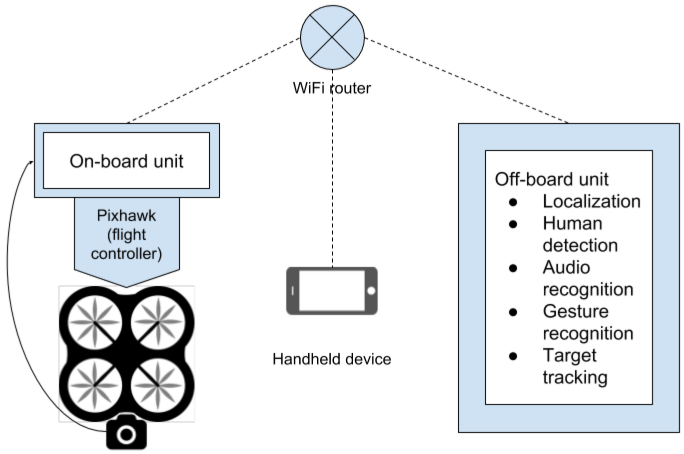

Figure 3. Communication subsystem setup

Threat Avoidance
Threat avoidance consists of detection of potential obstacles and path generation to avoid the obstacles. Path generation is accomplished by using potential fields method as explained in Navigation. Obstacle detection is accomplished using the combination of various sensors. A forward-facing stereo vision camera is used to detect frontal obstacles and provide their positions. Various strategies for obstacle detection in the remaining directions are:

- Monocular cameras is mounted in the other direction to detect any obstacles. CNN-based architecture is used for obtaining a bounding box for the obstacle. The obstacle can also be detected by obtaining velocity field for the image followed by subsequent thresholding. The position of the obstacles is determined using consecutive frame comparison and size expansion algorithms.
- Use of conical ultrasonic or infrared sensors for obstacle detection and obtaining their distance . Further the location and orientation of the sensor is used to get the position of the obstacle with respect to the quad-copter.

Communications
The mission requires us to coordinate among four quad-copters to enable swarm interaction. On top of that, the system has been designed such that all mission algorithms are run on an off-board computational unit. Hence all the four quad-copters need to transmit data to the off-board unit for processing and receive the commands in real-time. The bare-bones setup of the communications subsystem is shown in Fig.3. A 802.11 WiFi is used for all wireless communications for long range and easy setup with ROS. The main components are:

- A WiFi router with directional antenna.
- Four on-board computational units, one mounted on each aerial vehicle equipped with a WiFi transceiver module.
- An off-board computational unit equipped with a WiFi transceiver module.
- A android enabled hand-held device is to be carried by the player in the arena. It is connected to the same WiFi network

The *theora* ROS topic has been used for subscribing to all video transmissions on the off-board unit. This topic compresses and sends the video feed, so that the processing can happen in near real-time. The uncompressed ROS topic has a latency of 1.5 seconds per frame which is unacceptable for the purpose.

Power Management System

Two 4s Li-Po batteries with a total power capacity of 12 Ah are used to power all the components of the quad-copter. The selection of 4s was dictated by the high voltage requirement of the motors. The capacity was selected considering certain characteristics like the time of flight and the weight of the quad-copter.
As 2 batteries are used in parallel, a simple precaution taken, is charging them to the same voltage or fully so as to avoid unnecessary losses due to difference in initial EMF's. There are 2 power lines, one is a 16V main line to power the motors and on-board computational unit whereas the other is 5V auxiliary line to power all the other devices like LiDAR, PX4Flow, etc. The 5V line is maintained using DC-DC Step down converter.

OPERATIONS

Flight Preparations

Before any flight, the following checklist is completed:
Hardware:

1. Physical condition of the quad-copter is checked: No loose parts and objects in danger of being hit by the propeller.
2. Electrical systems are looked over for potential breaks in the wiring.
3. Battery voltage is checked.
4. RC transmitter is switched on.
5. Quad-copter is powered up.
6. Transmitter is readied to terminate flight via kill switch.

Software:

1. WiFi connection is established between on-board computational unit and router.
2. Autopilot is connected via MAVROS and roslaunch file is run to execute ROS nodes
3. Pilot stands by for manual override
4. Autopilot software takes control of the quad-copter
5. Login and initialize autonomous mode

Man/Machine Interface

Man-machine interface(MMI) is a crucial challenge in Mission 8. The main components of MMI are:

- **Vision module**: consists of human detection and tracking and commanding the quad-copter via gesture recognition.
- **Audio module**: consists of audio command recognition.

a. Human Detection *b. Gesture Detection*

Figure 4. Results

Vision Module
For human detection and tracking, a deep learning based approach has been applied. This method uses a novel model architecture that performs 3D convolution and is computationally less expensive. Point cloud approach has been used to find the distance of person from the aerial vehicle.
For gesture-based communication, image processing techniques have been applied on the results of human detection to narrow down the field of search as shown in fig.4a. On this result, differentiation of specific instructions is done using image manipulations on contours and removing background noise as shown in Fig.4b. In order to optimize the output, it is triggered by audio and information is extracted only at that instant, thus increasing its reliability.

Audio module
For audio-based communication, as our environment would be too noisy, we have opted for spectral subtraction approach to filter out the noise from the audio component of MMI. Since the mission allows us to use a hand-held device for recording audio samples we have used the popular CMU PocketSphinx system based on [4] for recognizing the audio commands on the device and then publish the results via ROS. The system has been implemented on Android using the rosjava template provided by [5].

RISK REDUCTION

Significant consideration has been given to risk reduction and crash worthiness in designing the quad-copter. All the components including are guarded in casings and guards that protect them from the mechanical damage due to potential collisions. Carbon fiber propeller guards have also been employed in the quad-copter.

Vehicle Status

Shock/Vibration Isolation

Shocks and vibrations are bane for any mechanical system. They reduce component life and affects sensor accuracy. To tackle the problem, shock absorbing dampers have been used to mount the flight controller. Also, the landing gear is capable to absorb sudden impulse as a result of hard impact on the ground. The motors are mounted on a carbon fiber frame, which has been tightly attached using lock nuts. This aids to reduce vibrations due to loosely attached parts.

Unbalanced rotating propellers produce significant vibrations that propagate throughout the quad-copter. Hence propeller balancing tools are used to balance the propellers before every flight.

EMI/RFI Solutions

The most EMI sensitive device on the quad-copter is the flight controller as it has a magnetometer. The external compass is mounted 15 cm away from the platform. Also, as a precaution, the ESCs and wires are rerouted, so that its effect on the results of the IMU data is reduced as far as possible. An EMI filter has also been incorporated in the electronics for redundancy.

Safety

The propellers rotate at a very high angular velocity, so it is of utmost importance to have them protected. Lightweight yet strong propeller guards using carbon fiber rods and 3D printed structures have been designed. The guards are covered by a lightweight nylon net which further adds to the safety. AnSys simulations were performed to make sure that the guards are fully man safe.

As the flight is expected to be fully autonomous, a kill switch is built and used as the last resort. The kill switch has the sole function of directly killing the power to all the motors upon receiving a trigger. Its control circuit includes a micro-controller to receive signal from the receiver and produce the required trigger pulse. This trigger pulse is used to switch a MOSFET ON and OFF in the power circuit, which supplies power to all the motors.

Modeling and Simulation

AnSys simulations have been performed on the quad-copter frame and propeller guards to check their durability when subjected to in-flight conditions. Simulation results as illustrated in Fig.5 showed a 10 kg static load bearing capacity for the propeller guards which is sufficient for the purpose.

The control systems of the quad-copter is simulated using a PX4 flight stack simulator implemented in Gazebo as shown in Fig.6. Simulations are performed prior to actual test to check for stability and proper functionality of the system. The competition arena has been incorporated in the simulator and mock sensors are used for system simulation. Static walls are used for threat detection and avoidance maneuvers.

Testing

All testings of the aerial robot and associated software subsystems are done in a controlled indoor testing environment which includes a sample optically sensitive floor pattern. Safety harness is also tied to the quad-copter while testing various software algorithms.

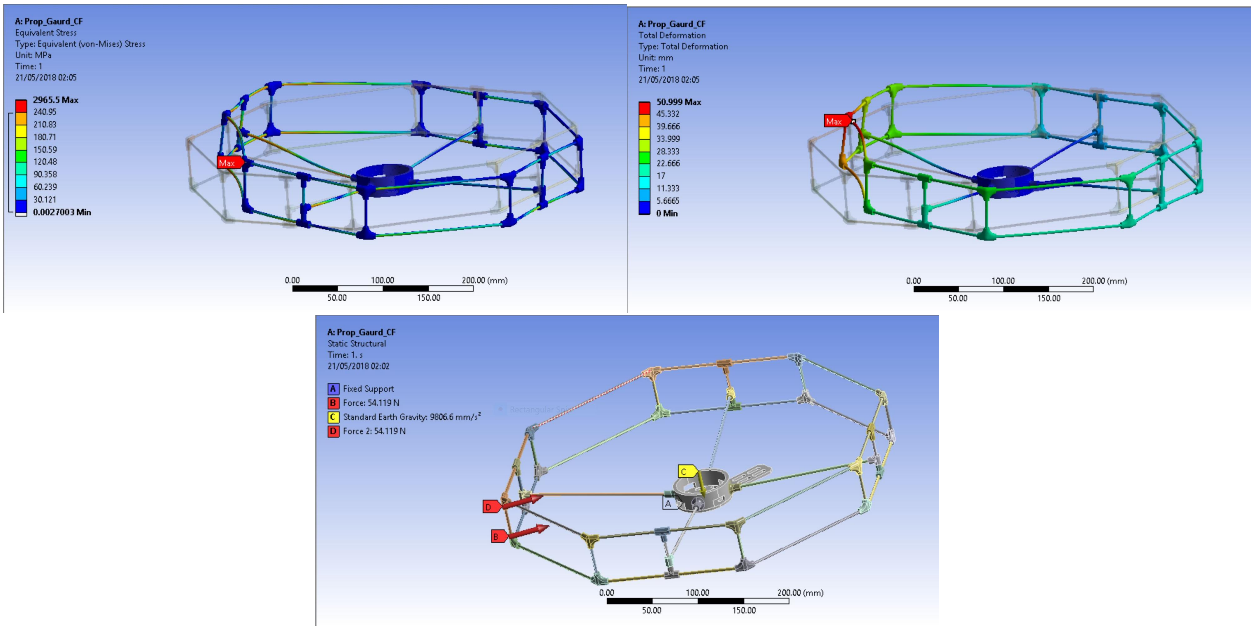

Figure 5. ANSYS simulation results

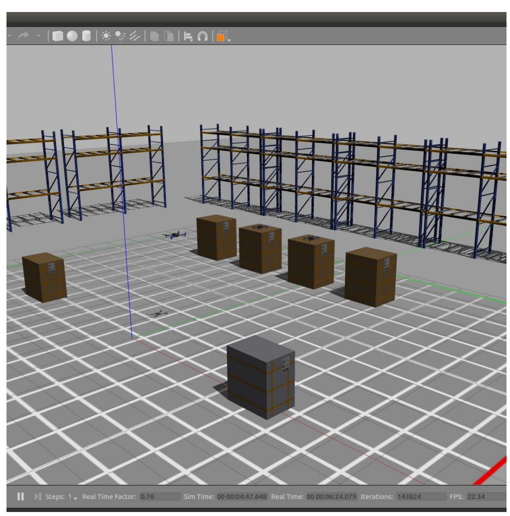

Figure 6. Gazebo simulation

At each stage of development, thorough computational simulations are performed before engaging in full-scale flight testing. Standardized tests were developed for individual subsystems of the quad-copter to allow easy evaluation of changes. Larger integration tests allow multiple subsystems of the quad-copter to be tested in simulations with a variety of different consecutive maneuvers. These same tests are then run on the real vehicle to validate the controller designs on physical hardware.

The stability augmentation system is tested in an outdoor environment to perform maneuvers and check stability. These tests form the basis for various navigation parameters. This also allow to confirm that the battery has sufficient capacity to sustain flight for more than eight minutes.

When testing custom-designed mechanical components of the vehicle, structural analysis is performed to verify mechanical stability of the components.

CONCLUSION

The aerial vehicle designed and developed by the team is structurally sound and capable of sustaining stable flight long enough to complete the mission statement. Various software techniques ranging from potential fields for obstacle avoidance, to a Convolutional Neural Network based model for human, and object detection have been employed. The algorithms have been tested and have achieved fairly high accuracy on the testing platforms, both in simulations and in real-world. Further work is required to ensure that all four quad-copters can accomplish the given mission statement in unison.

REFERENCES

[1] Joseph Redmon. Yolo: Real-time object detection.

[2] Ross Girshick. R-cnn: Regions with convolutional neural network features.

[3] Ross Girshick. Fast r-cnn.

[4] David Huggins-Daines, Mohit Kumar, Arthur Chan, Alan W Black, Mosur Ravishankar, and Alexander I Rudnicky. Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 1, pages I–I. IEEE, 2006.

[5] ollide. A small android studio project containing all ros dependencies, an example node and a proguard configuration.