Overview

Acme Robotics is building a next generation mobile robot to be used on sports fields to detect and track players and provide analytics data about them to be utilized for broadcasting. A swarm of robots will be deployed on the field.

The mobile robot will have a monocular camera placed on it to receive live video feed of the field. A perception module on the robot will receive the video feed and will label each player and store their location with respect to the robot. The output of the module will be the labels and location of the players. This will be fed to a path planner, so that the mobile robot can track a player to ensure that the player is in its field of vision.

- We plan to use Machine Learning, specifically Convolution Neural Network Algorithms to detect, classify and track objects from a video stream.
- In order to perform deep neural computation we are using open sourced computer vision libraries such as openCV and YOLO with C++
- For computation we assume that Nvidia Jetson Nano 2GB chipset for deep learning computation is used.
- The targets are assumed to be 180cm in height, which will be used to calculate distance from the robot in the forward direction.
- Assuming MX219-160 Camera with 3280 × 2464 Resolution, 8 Megapixels and 160° FOV for video feed.

- YOLO: The YOLO algorithm uses convolutional neural networks (CNN) to detect multi objects in real-time. This algorithm uses a single forward propagation through a neural network to detect objects in a single run.

- OpeCv: An open-source library for computer vision, machine learning, and image processing.

- The live stream of video data path from robot's monocular camera is fed to a c++ parser to interface it with openCv libraries. Using OpenCV we capture each Frame of the video and convert it into a 4D Blob, This blob is fed to a convolutional neural network to detect, classify and track the object. Human detection and tracking is done through bounding box regression and a probabilistic estimation of confidence score. If the confidence score is less than the threshold we are eliminating the detected box. Similarly if the confidence score matches the expectation, we are displaying the detected object.

Deliverables

- **Timeline:** The estimated time to deliver perception module is within 16 days i.e on or before 22nd Oct 2021
- **Product Backlog:** we are following Agile iterative process methodologies for product delivery with a burn down chart.
- **Possible Deliverables:**
  - Perception module which outputs detected labels of humans and its location from robots frame
  - Quality Assurance of the module

- **Testing:** Google's Open Images Dataset V6+ will be used for testing and quality assurance. The dataset contains labelled images of humans with the information about bounding boxes provided in it. For real time testing, a laptop and webcam will be used and the output will be monitored to ensure proper functioning.