ENPM700 Software Development for Robotics

Midterm Phase 0

Proposal: Human Obstacle Detection and Tracking Module

Abstract: In a Social Robot system, the detection and tracking of humans is an essential component for a safe and efficient operation in environments where human-robot interaction is common. A robot can precisely plan to adjust its trajectory to avoid collision and maintain safe distance from humans by identifying the humans and tracking them. We propose a Human Obstacle Detection and Tracking Module using YOLOv8, tailored for a robotic system intended for navigation in human-populated environments at Acme Robotics. This module will enable the robot to detect and track multiple humans in real-time using a monocular camera. The module will output location in a robot reference frame.

Team Members:

- Soundery Varagur Venugopal (Driver)
- Anirudh Swarankar (Design Keeper)
- Amogha Thalihalla Sunil (Navigator)

Interface with Other Components*:*

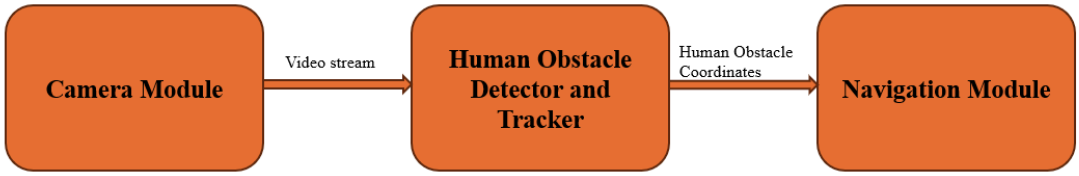

Fig 1: Interfaces

Assumptions

The camera is mounted at a fixed height to provide a consistent reference frame. Further the robot has basic navigation capabilities and can interpret location data.

Design Process

The project will follow the AIP Process and implement the software using Test Driven Development. The software design is captured using UML Class diagram and Activity diagram.

Development Process

We plan to utilize YOLOv8 for human detection and OpenCV for video processing and later perform unit tests in a controlled environment. We shall also maintain comprehensive documentation throughout development, including design choices and interfaces. Implement unit tests and integration tests to verify functionality.
The implementation will be done in C++ (with CMake build), libraries including OpenCV and eigen will be used.

YOLOv8 (You Only Look Once) [1][2] algorithm is a real time object detection and tracking model which is light weight and fast. It works in a way that it divides each input image into smaller region and predicts bounding boxes and confidence for them, the confidence measures the degree by which the model is sure about the object prediction in the bounding box. Further it tracks the frame wise detections and develops unique descriptors to each detection and tracks them as they move around through the frames (here frames are from time 0 to time n since we are considering a monocular camera). Each tracked object has a bounding box and an associated confidence of the prediction. Finally Coordinate transformation will be performed to convert pixel coordinates to coordinates in robot's frame in 3d.

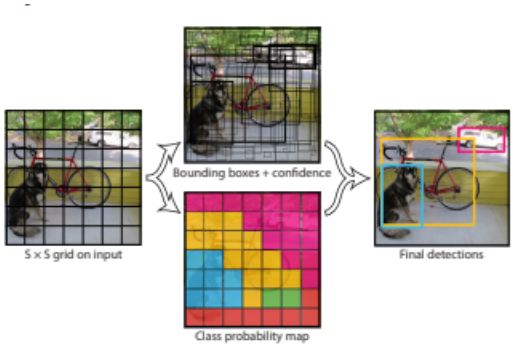

Fig 2: YOLO model

Risks and Mitigation

There are chances of potential performance issues with real time processing, if the model is not well optimized. There can also be false positives for unseen data while using the pretrained yolo model.

Final Deliverables to Acme

A fully functional Human Obstacle Detection and Tracking Module along with comprehensive documentation that includes design specifications, source code and developer documentation.

References

[1] You Only Look Once: Unified, Real-Time Object Detection, Joseph Redmon and Santosh Divvala and Ross Girshick and Ali Farhadi, 2016,1506.02640

[2] Real-Time Flying Object Detection with YOLOv8, Dillon Reis and Jordan Kupec and Jacqueline Hong and Ahmad Daoudi, 2024, 2305.09972

[3] https://docs.opencv.org/4.x/da/d9d/tutorial_dnn_yolo.html