PERCEPTION MODULE TO DETECT AND TRACK HUMANS

Guru Nandhan A D P
M. Eng Robotics
University of Maryland, College Park
guruadp@umd.edu
(Driver)

Vignesh R R
M. Eng Robotics
University of Maryland, College Park
rr94@umd.edu
(Navigator)

Dhanush Babu Allam
M. Eng Robotics
University of Maryland, College Park
dallam@umd.edu
(Design Keeper)

I. INTRODUCTION

Over the past decade, there has been consistent work and improvement in the field of Robotics and Computer vision for providing more robust autonomous solutions in various domains like Automotive, Manufacturing and Medicine.

The primary aim of such a solution starts with monitoring the environment and detecting objects/obstacles with reference to the robot and thus providing an optimal trajectory for its movement without colliding with the obstacle. Humans are the objects in most applications for mobile robots, thus human detection and tracking is of primary importance.Human detection and tracking are performed on computer vision systems for locating and following people in the video imagery. In most cases, a camera or lidar is used as a sensor to detect the objects and is proven to be highly efficient.

Human detection is the task of locating all instances of human beings present in an image (one frame of the video), and it has been most widely accomplished by searching all locations in the image, at all possible scales, and comparing a small area at each location with known templates or patterns of people (or their attributes) as reference. The job is not done yet, as the sensors provide dynamic data that requires constant detection of existing and new objects, and also tracking the movement of the objects with reference to the previous frame.

Human tracking is the process of temporally associating the human detections within a video sequence with an ID and then tracking the changes to generate persistent paths, or trajectories, of the people. Human detection and tracking are generally considered two important processes for tasks such as primary action recognition, dynamic scene analysis and decision making.

II. DELIVERABLES

1) A software that detects and tracks humans for Acme Robotics in C++
2) A proposal with sprint, risks and mitigation and UML diagram(s) of classes and interfaces.
3) A proposal with sprint, risks and mitigation and UML diagram(s) of classes and interfaces.
4) GitHub repository with informative commit history and README.
5) Code stubs and unit tests integrated with Travis and code coverage with Coveralls
6) Developer-level documentation

III. SOFTWARE PROJECT MANAGEMENT PLAN

We will be following the Agile Iterative Process (AIP), wherein each module will be incrementally developed in each iteration to arrive at the final build. Every iteration is based on sprints that have the following actions : Planning, Development (Build), Testing and Review. On successful completion of one iteration, the next begins. Meanwhile, in the development process, stubs will be used for the development of multiple modules independently. The everyday progress will be discussed over scrum meetings to check for product backlog, implementation status and blockers (if any).

IV. METHODOLOGY

We have planned to use a pre-trained yolov5 model for the detection. Out of the 80 classes which can be detected by the yolo model, the humans are

filtered that have a certain level of confidence as percentage. The confidence level is the likelihood of the proper detection based on the comparison of the image with the reference patterns mentioned above. If the confidence levels are above the required threshold, then the object is classified according to the detection. Each detected human will be assigned with a unique identifier and tracked. In the final stage, the camera pixel positions are transformed into the robot's frame.

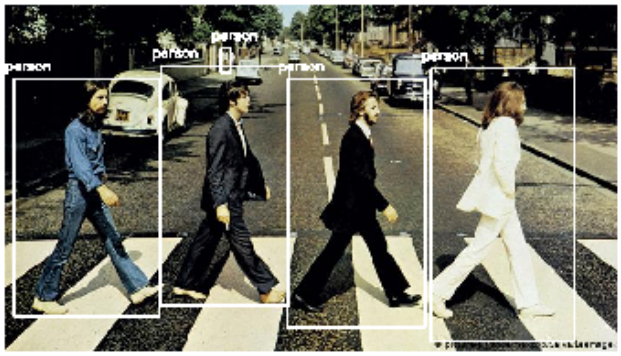

Fig. 1. Example of YOLO detecting humans

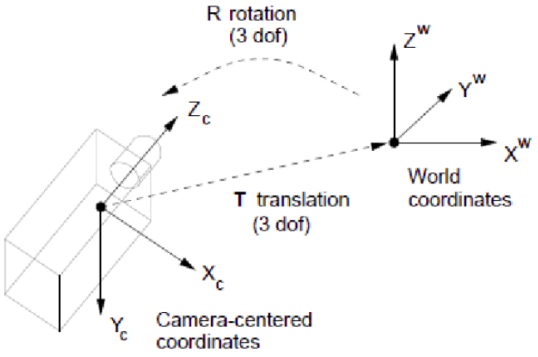

Fig. 2. Camera coordinates to Robot coordinates.

V. Assumptions

1) Ground is flat
2) Camera is not tilted
3) No noise or distortion
4) Average human height is 175.25 cms
5) Stationary camera mounted on the mobile robot
6) The same ID can be assigned to a different human if the previous holder of the same ID left the frame before entry of the new person.

VI. Challenges and Risks

Live video input will have dynamic data that may have occlusions where multiple humans will be closer to detect independently and the confidence levels will be lower. Calculating the distance between the camera and detected humans. Errors in detecting the object due to the environmental conditions such as poor illumination of the object.

References

[1] https://github.com/ultralytics/yolov5
[2] https://docs.ultralytics.com/tutorials/torchscript-onnx-coreml-export/
[3] https://medium.com/@luanaebio/detecting-people-with-yolo-and-opencv-5c1f9bc6a810