Acme Robotics Human Detector-Tracker

SHAIL SHAH, JAY PRAJAPATI AND ANUKRITI SINGH

I. INTRODUCTION

The purpose of human detection is to find the location of every person in each video image, while producing as few false detections as possible. The problem of detecting, localizing, and tracking human presence has been a widely studied problem due to its potential military, safety, security, and entertainment applications. Human detection is the task of finding location of all instances of human beings present in an image, and we are going to accomplish it by searching all locations in the image and finding the location of human with respect to the robotic frame. Human tracking is the process of relating time with associating the human detections within a video sequence to generate persistent paths, or trajectories, of the people.

II. DEFINITIONS AND ACRONYMS

Computer Vision: Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images and videos.

Detection: Detection is a computer vision technique for locating instances of objects in images or videos.

Tracking: Visual tracking is to estimate the future position of a visual target that was initialized without the availability of the rest of the video.

Segmentation: It is the process of dividing an image into different regions based on the characteristics of pixels to identify objects or boundaries to simplify an image and more efficiently analyze it.

Neural networks: A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates.

Features: The representations that are learned by the previous network to extract the interesting features from new samples.

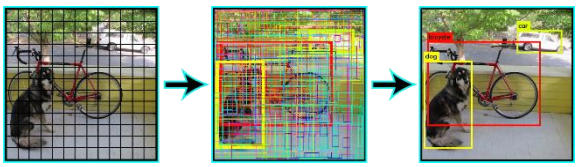
III PROCESS

Acquiring image

Out robot will do real-time object detection and tracking, hence we use a camera to capture image or video on which the image processing will be done. It will also have auto calibration mode due to which the use does not need to calibrate the camera for capturing.

Image processing and Detection algorithm

We use YOLO (You Only Look Once) real-time object detection system. YOLO is a state-of-the-art, real-time object detection system. It is extremely fast and accurate and has distinct versions available. We will use a pre-trained YOLO model. We apply a single neural network to the full image. The YOLO network divides the image into regions and predicts bounding boxes and their probabilities for each region. The algorithm weighs bounding boxes and choses them according to probabilities. We use MS-COCO dataset, a large-scale object detection, segmentation captioning dataset available online. It contains 164k images divided into training and testing



Tracking algorithm

The tracker module will have detected human bounding box and its confidence value. Motion predictor predicts subsequent positions of each tracked object. Feature predictions are calculated using similarity scores between detection couplets. OpenCV based object tracker like KCF tracker is used.

Relative position mapping

This module provides output as location of human in robot reference frame. This helps to know where the object in real world and there are several applications based on this.

Homogeneous Transformation Matrices

- 3x3 Rotation Matrix $T_1 = \begin{bmatrix} C1 & -S1 & 0 \\ S1 & C1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
- 3x1 Displacement Vector $R_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$
- 4x4 Homogeneous Matrix $T_H = \left[\begin{array}{ccc|c} C1 & -S1 & 0 & x_1 \\ S1 & C1 & 0 & y_1 \\ 0 & 0 & 1 & z_1 \\ \hline 0 & 0 & 0 & 1 \end{array}\right]$

ORGANIZATION

Design Keeper	Driver	Navigator
Anukriti Singh	Shail Shah	Jay Prajapati

TECHNOLOGIES

C++14 and above	OpenCV library
YOLO algorithm	Auto-Calibrated Camera

MANAGEMENT

In our concept of software development and implementation for ACME Robotics, we will be using upfront class of models called ‘designs’. This will cover main features/characteristics of the software that we will be building. We will integrate our class of model with Agile Iterative software Process (AIP) using Test Driven Development approach. As Agile Iterative Process involves decision making by consensus, this would result in a holistic consideration of all software development procedure and curb any misinterpretation of client requirements.

RISK IDENTIFICATION AND MITIGATION

One of the risks involved in the perception modeling would be to keep the identity of identified human discrete.

The perception module must be in line with ACME Robotics’ product dependencies to be able to ease the integration and implementation of formulated algorithm.

REFERENCES

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

[2] Real-Time Human Detection, Tracking, and Verification in Uncontrolled Camera Motion Environments Mohamed, Hussein Wael Abd-Almageed Yang Ran Larry Davis Institute for Advanced Computer Studies University of Maryland.

[3] https://appsilon.com/object-detection-yolo-algorithm/

[4] Zhao Z.Q., Zheng P., Xu S.T., Wu X. Object detection with deep learning: A review IEEE transactions on neural networks and learning systems, 30 (11) (2019), pp. 3212-3232.