Computer vision: Training a 3 DOF robotic arm to detect, classify, and pick up objects

Sean Morton, Northwestern University
Anirudh Adiraju, Vernon Hills High School

Mechanistic Data Science - Northwestern University - Summer 2021 REU
Professors W. Liu / M. Fleming / Z. Gan, Northwestern U.

9 July, 2021

Objective

The goal of this project is to create a robotic arm, for low cost, that can detect and classify objects in real space. Once a computer or microcontroller has processed the visual info, it will send commands to the robot: to move the arm to the position of an object, drop down, and pick up the object.

Approach

Methods:

- Robot arm will be a 3DOF, 3D-printed SCARA arm
- Robotic Operating System (ROS) will be interfaced with Arduino to control the robot
- OpenCV will be used to record video from 3 cameras: front/top/side views
- Tensorflow and/or Pytorch modules in Python will be used to train convolutional neural networks (C-NN) for the sake of real-time object detection

System and design

Applications of system: allows for hands-free control of a robot arm, controlled only by commands of which item to pick up: “banana”, “apple”, “orange”. Could be highly applicable in medical settings, i.e. to give commands to an arm to pick up a scalpel. Could also be applicable to people who cannot pick up items by themselves--can instruct the robot to pick up a mug and bring it to their mouth, for example

Limitations of system: Convolutional neural network will only know how to classify objects that it’s previously been trained on. The robot arm, meanwhile, needs a method to pick up irregular objects, like the handle of a coffee cup--constructing a 3D mesh of the objects in space was one suggestion to gather data to help the robot decide how to pick items up.

Multimodal data generation and collection

Object position and class data will be collected from three cameras: on a standard front, top, and right view of the table where the robot arm is handling objects. The table itself will be white with black gridlines on a mat. Fiducial Markers will be used to translate position in images to position in the real world. With these markers, the code can collect data on position in the image, then convert to position in the real world where the robot arm should move.

Feature Engineering + Dimensional Reduction

Features of input pictures:

Resolution of image	Location of bounding boxes for objects in xy, yz, xz planes	Background noise
Size (width, height) of image	Classification of objects	Reference geometry (fiducials)
Number of objects per image	Speed of each object's motion	

Relationships between features:

- Bounding boxes from the three cameras will have some redundancy (one camera has x+y info; another has y+z; another x+z)
- Classification for objects should be consistent between different cameras, with different levels of certainty (ex. 95% certainty when looking head-on at a water bottle, versus only 40% certainty when looking from above)
- Number of objects detected should be consistent between different cameras, unless one object blocks another

Dimensional reduction will be used most prominently in the reduction of the (x,y,z) coordinates of the bounding boxes of each object. The three cameras will give coordinates (x_a, y_a), (x_b, y_b), and (x_c, y_c) that can be reduced to coordinates in 3 dimensions (x,y,z).

Regression + classification

A Convolutional Neural Network will be used to train the object detection system. Classification of objects and their bounding boxes will work using pre-trained models in Tensorflow and Pytorch.

Summary

We expect to be able to have a working robotic arm with object detection enabled by the end of five weeks. Whether or not we will (in 5 weeks) be able to convert the positions of bounding boxes in an image to position in the real world is yet to be determined. More research is needed to determine the feasibility of using fiducial markers as a way to get a basis for position in the real world.

Possible extensions of the project: if given the opportunity to work more on this project in the future, one direction we would love to take this project is voice commands. In other words, to be able to control the robot by saying "Ok, robot: pick up the cup." For the robot to be able to "see" objects with computer vision, and to "hear" verbal commands using audio recognition models, would be the ultimate goal for an easily controllable robot.