

Introduction

Recent decades witness a great deal of developments in Computer Vision, Natural Language Processing, Artificial Intelligence and Robotics. As a result, it attracts a lot of research interests to apply these cutting edge technology for improving daily life. Home Service Robot(HSR) researches are some of the examples. Among the problems HSR researchers are confronting with, interaction interface between Human and Robot is one of the major one. Fundamental goal of this proposal is to make it easier for Human and Robot interacting with each other. By combining techniques in Computer Vision and Artificial Intelligence, it's possible to build a demo-based interaction interface.

Robot could be of great help to human being's daily life, especially with some tedious and time-consuming tasks, such as cooking, laundry, cleaning floor, washing dishes. Furthermore, robot can replace human beings on some jobs, such as goods placement in supermarket, servant in restaurants. However, there are a few technique obstacles to deal with before robot could be of real help. One of the most difficult problem is the abundance of requirements by different users. Although it's easy to design specified robot on some jobs such as floor cleaning, it's still challenging, if not impossible, to build a robot that is capable to fulfill all the required tasks. One of the possibilities is to build a robot that is able to learn a task from the user. This could be paralleled with human being. Human beings are born with none of the knowledge about this world, but they are able to learn from parents. One of the most important ways to teach infants is by demonstration. Thus, this proposal aims to improve robots' capability by Learning from Demonstration(LfD), which is one of the possible way for users to configure robots for different tasks.

This proposal is arranged as follows. The project is summarized in the first following section to . Then the Intellectual Merit section is followed to indicate the contributions that this proposal will bring to related researches. Last but not least, potential applications of this framework is listed.

Project Summary

There are two phases in this framework. The first phase is learning phase and the second phase is action phase.

Taking camera frames as input, the system learn real world rules from demonstration by a supervisor. This supervisor is not special trained technicians, but regular common users. What the supervisor needs to do is to demonstrate what he or she want the robot to know, such as the rule "drain the cup first before refill new beverage". For the robot, the learning system will analysis the camera and voice instructions, learn the rule and action accordingly later. What is worth noting is that firstly, this demonstration could be done by anything without any special training; secondly, this kind of unconstrained rules exist everywhere, which makes it impossible to be preprogrammed into robot memory.

Based on the learned rules, the robot is able to make decision accordingly. Next time, for example, the user wants the robot to get a cup of orange juice, but the robot can only find a cup with apple juice. The robot will dump the apple juice first. Another thing needs to do is Action Plan. In some scenario, there are more than one solution to a task, and Action Plan is to choose to optimal action sequences to execute.

Project Description

In the learning phase, the frames is processed to identify different objects and the interactions between them. After that, a video semantic recognition will be done to understand the objects and interactions semantically. Then the semantics token will be feed into an AI Solver to update the rule database. In the video understanding process, following techniques are required.

- Camera Calibration
- Object Recognition
- Action Recognition
- Video Semantically Analysis

Other than video analysis, the Natural Language Processor is built to understand the voice instructions, whose output are also the semantically tokens that will be feed into an AI Solver.

The fundamental part of this proposal is the AI Solver. Both Video and language understanding take the AI solver as a reference to realize semantics analysis. A novel AI Solver will be created to integrate information from video and language understanding. This AI solver is an incremental AI solver, which is able to learn new rules and update world model accordingly. One thing worth noting for this solver is that the initial rule set and world model are not empty; on the other hand, some of the rules and world model have been programmed into the system, above which the system learn new rule and world model. For example, the solver will know what is the attributes of apple juice, before it learn “drain the apple juice first”. One of the problems this framework needs to address is to build this initial solver.

Broader Impacts

One of the most influential problems for robot to become popular is the inconvenient interaction between human beings and robot. Most of the current robot has either no interaction or obscure interaction with human being. With the demo-based interaction system, the robot could interaction with human being intuitively and naturally, which will boost robotics industry. Some humanity field in which robots have already play an important role, such as disabled and aged people assistance, will benefit from this. Furthermore, since this system makes it possible to train robot do more complex jobs, new demands for robot will be created as well.

Biographical Sketch: Yuncheng Li

Currently, I am a graduate student at University of Rochester , Department of Computer Science, pursuing PhD degree. My advisor is Prof. Jiebo Luo. Before coming to UR, I finished my bachelor degree at University of Science and Technology of China, Department of Electrical Engineering and Information Science.