Parsing an action sequence in DFT

David Lobato
Outline

- Introduction
  - The problem
  - Learning from Demonstration
- System overview
- Ongoing work
Introduction

The problem

- Understand actions performed by an agent (teacher)
- Store sequences of seen actions
- Vision-based
- Real time processing
- DNFs
Introduction
Learning from Demonstration

- A.k.a Learning by demonstration, Programming by demonstration, Learning from Observation,...
- Teach an agent based on demonstrated data
- How demonstrations are gathered:
  - Sensors on agent: Teleoperation, shadowing (mimic)
  - Sensors out of agent: Sensors on teacher, external observation
Introduction

Approach

- External observation through a Kinect device
- Transitive actions: Subject + Object + Action
- Actions inferred by observed cues
- Correspondence problem not approached: Actions + parameters
System overview

Setup

- Input comes from a Kinect: RGB + Depth = RGBD
- Extensive use of open source libraries: PointClouds and OpenCV
System overview
Complete architecture
System overview
Low level sensory acquisition

Raw input is processed by libTabletop:
- Segmentation
- Motion analysis (2D at the moment)
- Object classification
- Efficiency > 20 fps
- Open source, available at github
System overview

Low level sensory acquisition

Motion detection

![Motion detection diagram]

Approach

![Approach diagram]
System overview
Low level sensory acquisition
Integrated as a cedar plugin:
System overview
Selecting the object

Attention mechanism focus on the object that will likely receive an action:

- Prediction
- Based on motion cues
- Before object is occluded
System overview

Selecting the object
System overview

Action detectors

Implemented action detectors:
  - Reach
  - Grasp
  - Drop

Other detector in process:
  - Lift
  - Push
  - Stack

Combinations to create complex actions: \( \text{Grasp}(p) + \text{Drop}(p') = \text{Transport} \)
System overview
Action detectors: Reach

Conditions that mark initiation, failure or success:

- Col: Object is being approached
- CoF: Col + No motion and hand not on top of object
- CoS: Col + No motion and hand on top of object
Ongoing work

- More actions (lift, stack, push)
- Detect complex actions: grammars?
- Store actions and parameters as a sequence: serial order model
- Detect chunk of actions
Thank you for your attention!

dlbravo@ualg.pt

Support:
Portuguese Foundation for Science and Technology (LARSyS)
EU project NeuralDynamics, FP7-ICT-2009-6, PN: 270247