Reinforcement Learning on a Kinect-enabled Robot  
LU, Yongxi  
Supervisor: Prof. Bertram Shi

Overview
In this project, we explore reinforcement learning on a Kinect-enabled robot. Reinforcement learning algorithms enable a robot to learn its behaviors through interaction with the environment. With the Microsoft Kinect sensor, the robot can make decisions based on the depth information of the environment.

Background
- **The Kinect sensor**
  The Kinect sensor provides the depth information of the environment. The depth is the distance to the surrounding objects (the z-direction in xyz coordinate).

- **The Reinforcement Learning**
  Intuitively, it concerns how an agent (in our case the robot!) learns to optimize its long term reward by interacting with the environment and receiving short term rewards along the way. In other words, learning and control at the same time. Its output is an optimal policy, that is, a rule that assigns an action for each state of the environment.

Challenges
- **Perception**
  - Compression of the high dimensional data
  - Robust features (simple description of the complex data) estimation under noise

- **Learning/Planning**
  - Long term goal versus short term reward
  - More exploration versus better decisions

Methodology

### Design

**Task description**
- The robot should move as fast as it can
- The robot should bump into obstacles as infrequent as it can
- All should be done using Kinect’s depth information as the main input

### The design of learning/control

**The learning algorithm**
- State of the environment
- Reward from the environment
- Actions that can be taken

**Our task**
- Features extracted from the Kinect input
- Punish when hitting obstacles, reward when going forward
- Simple actions

### Feature Extraction

- **Point cloud Construction**
- **Plane Segmentation**
- **Filtering**
- **Occupancy grid construction**

### Results

**Real-time plane segmentation**
- An novel algorithm based on RANSAC is derived for feature extraction
- The new algorithm achieves good performance in speed and robustness
- The new algorithm can be integrated into the control framework easily

**Online policy learning**
- The learning algorithm converges to a policy that maximizes long-term reward
- Better way to design the reward function which reflects physical performance requirement is found important