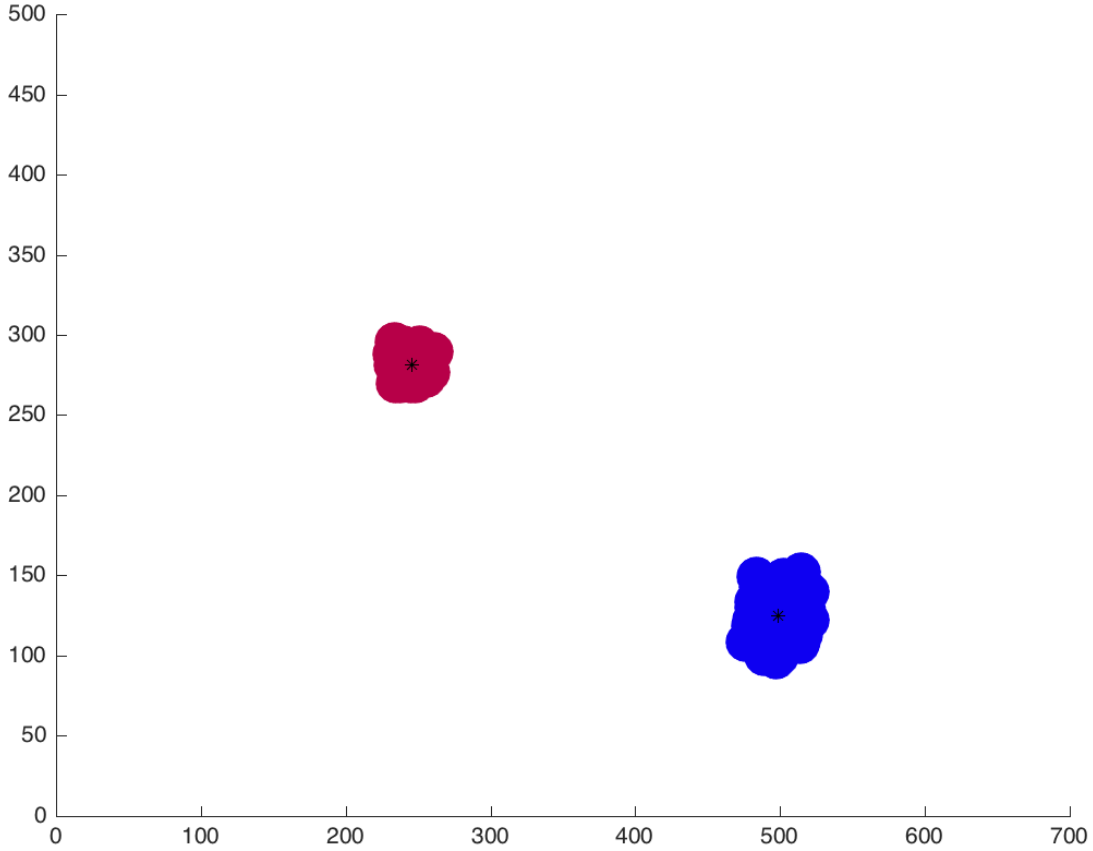
Optimal gaze control

THROUGH REINFORCEMENT LEARNING

What is it really?

A biologically inspired robotics model.

A general learning strategy for where to look to optimise the current task.


500
450
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700

Previous Work

Optimal gaze control thesis by Nunez Varela

Decision under uncertainty paper by Rashej Rao (2010)

Prerequisite

1. Particle Filters
2. Reinforcement Learning
3. Radial Basis Functions

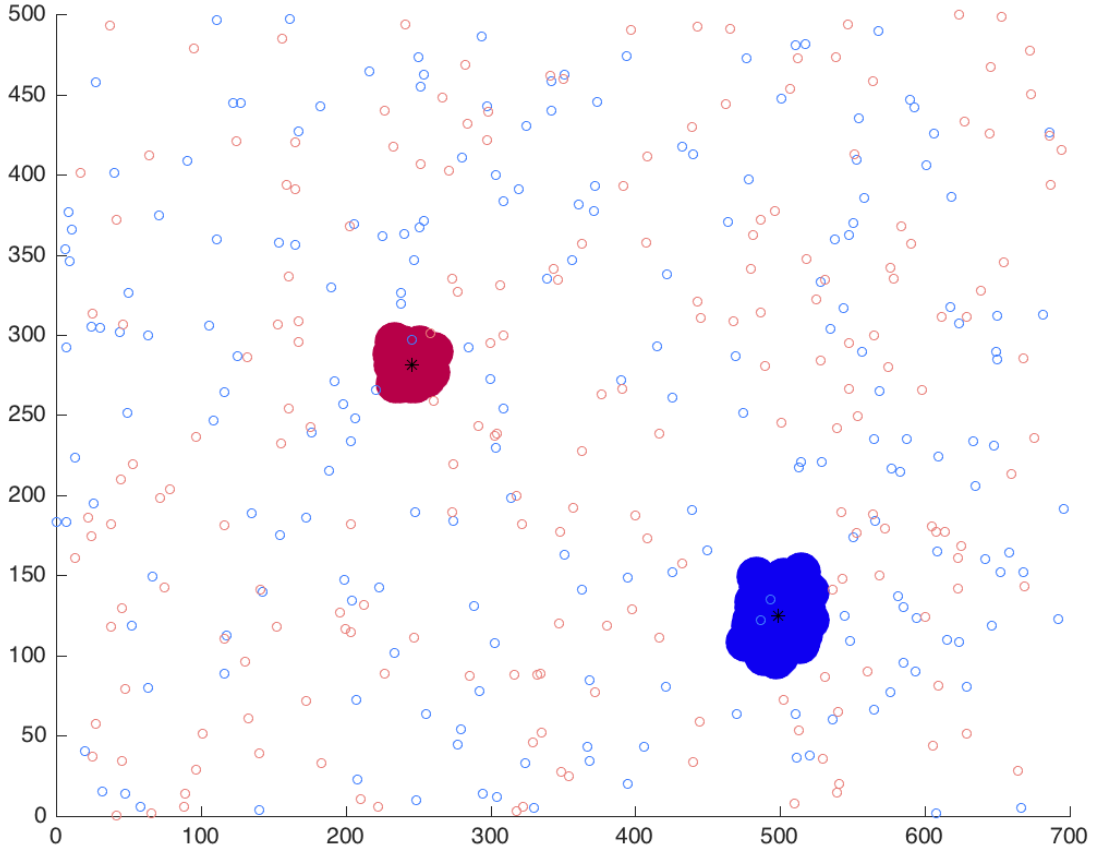
Prerequisite

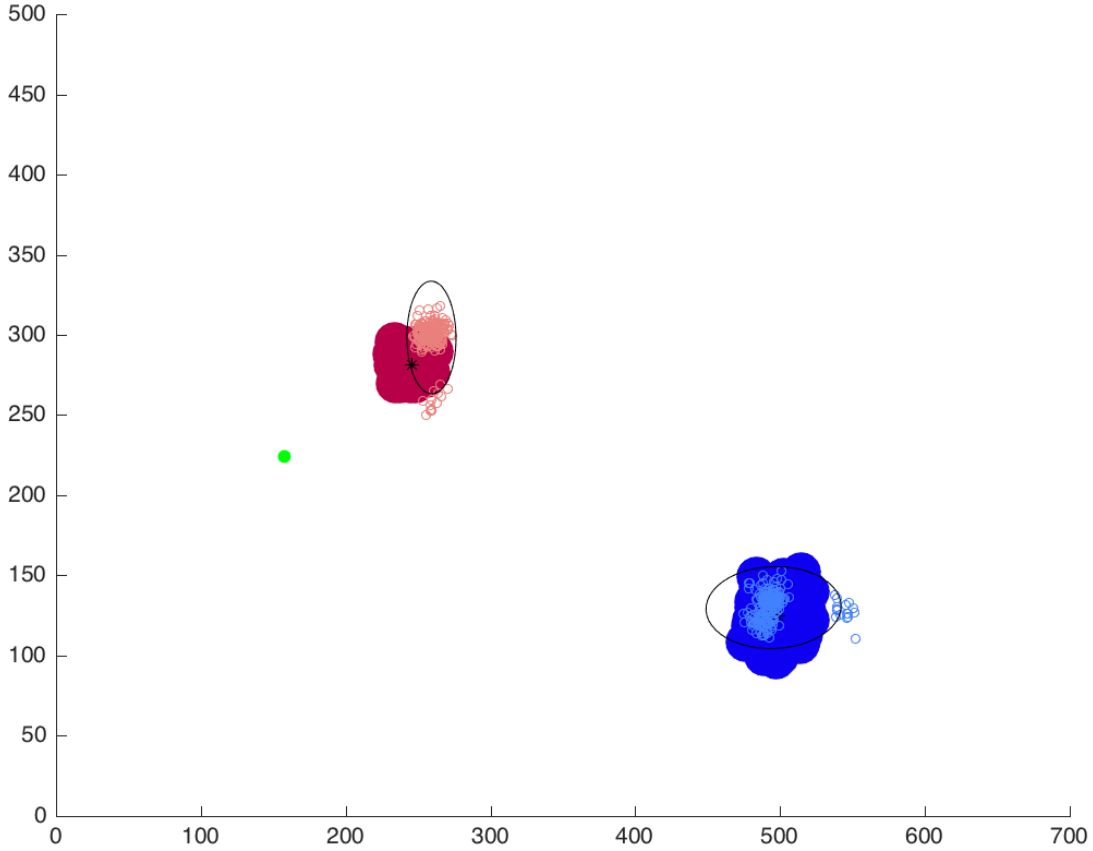
1. Particle Filters
2. Reinforcement Learning
3. Radial Basis Functions
4. Learning formula for where to optimally look

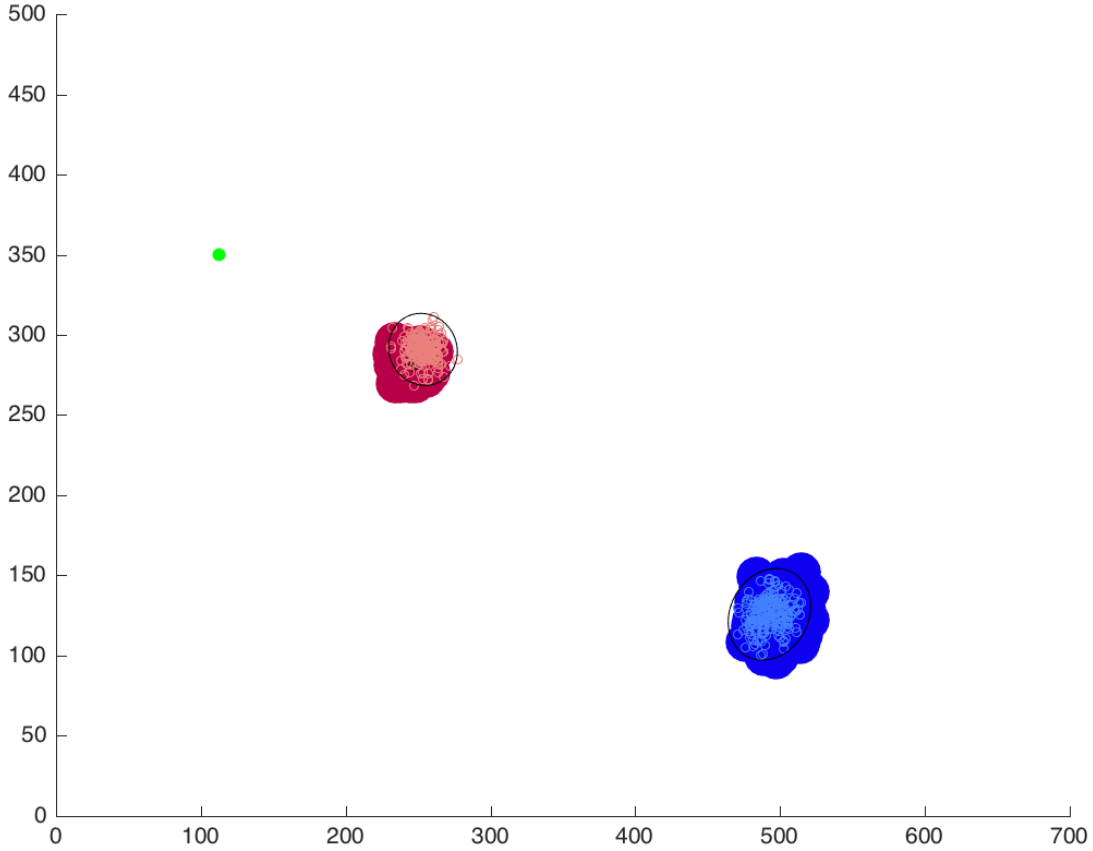
1. Particle Filter

Approximates the state of the system (i.e. object location).

Approximates the current uncertainty on the object location.


500
450
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700


500
450
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700


500
450
400
350
300
250
200
150
100
50
0
0
100
200
300
400
500
600
700

2. Reinforcement learning

Teaches robot task objective through positive and negative rewards

$$R = \begin{array}{c} \\ \text{State} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} & & & \text{Action} & & \\ 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} -1 & -1 & -1 & -1 & 0 & -1 \\ -1 & -1 & -1 & 0 & -1 & 100 \\ -1 & -1 & -1 & 0 & -1 & -1 \\ -1 & 0 & 0 & -1 & 0 & -1 \\ 0 & -1 & -1 & 0 & -1 & 100 \\ -1 & 0 & -1 & -1 & 0 & 100 \end{bmatrix} \end{array}$$

image source:
http://mnemstudio.org

Problems

Suitable for discrete space, but our space has continuous values.

What do we do ?

Problems

Suitable for discrete space, but our space has continuous values.

What do we do ?

Store the reward for state – action pairs in a neural network.

Radial Basis Functions

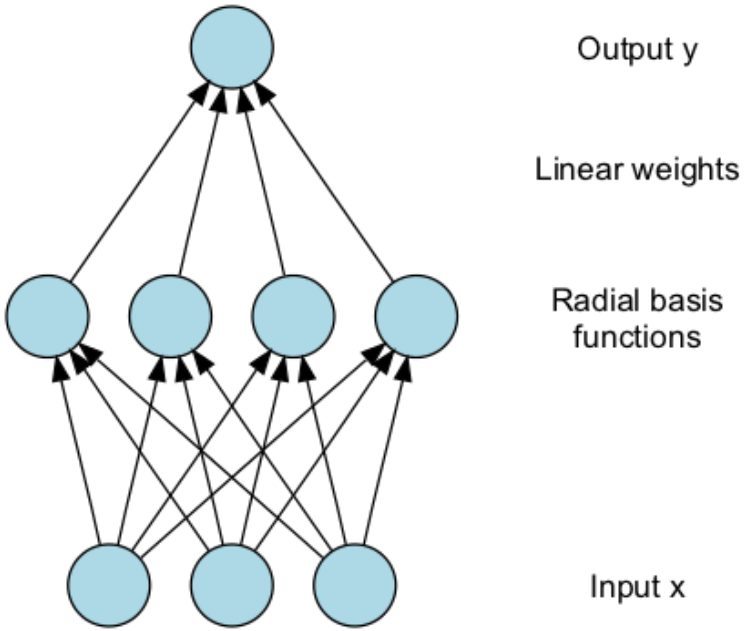


image source:
http://en.wikipedia.org/

Rashej Rao

Radial Basis Functions are used to store the reward and action.

Rashej Rao

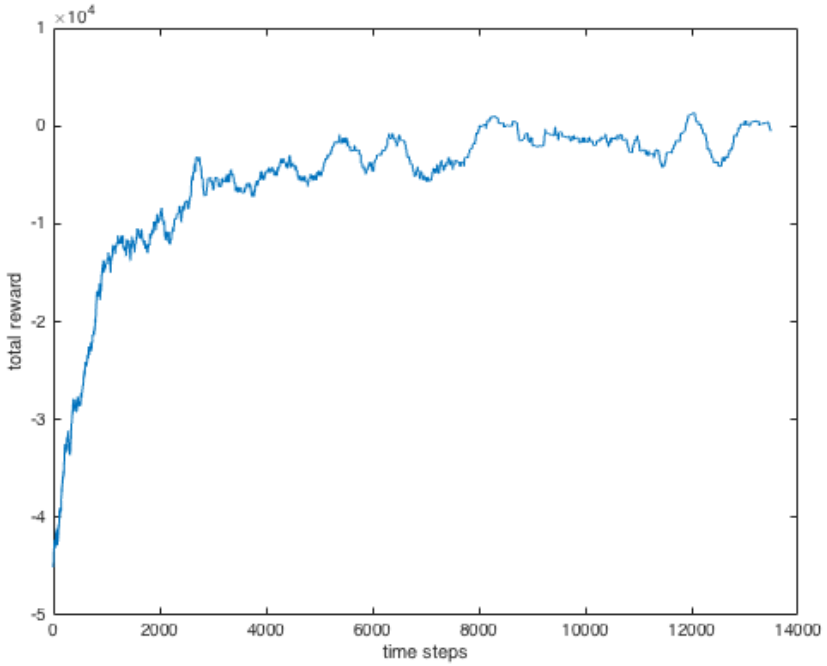
Radial Basis Functions are used to store the reward and action.

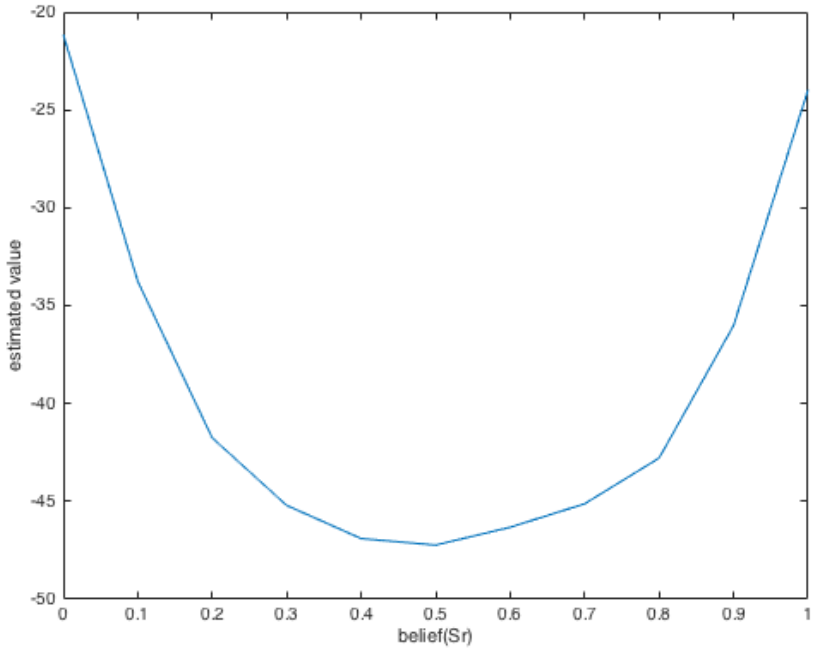
Network input is the belief space.

For Rao, it is a state $\mathbf{b_s}$ = [p(left) p(right)]

In our case it is the eigen value and reward value of each target.

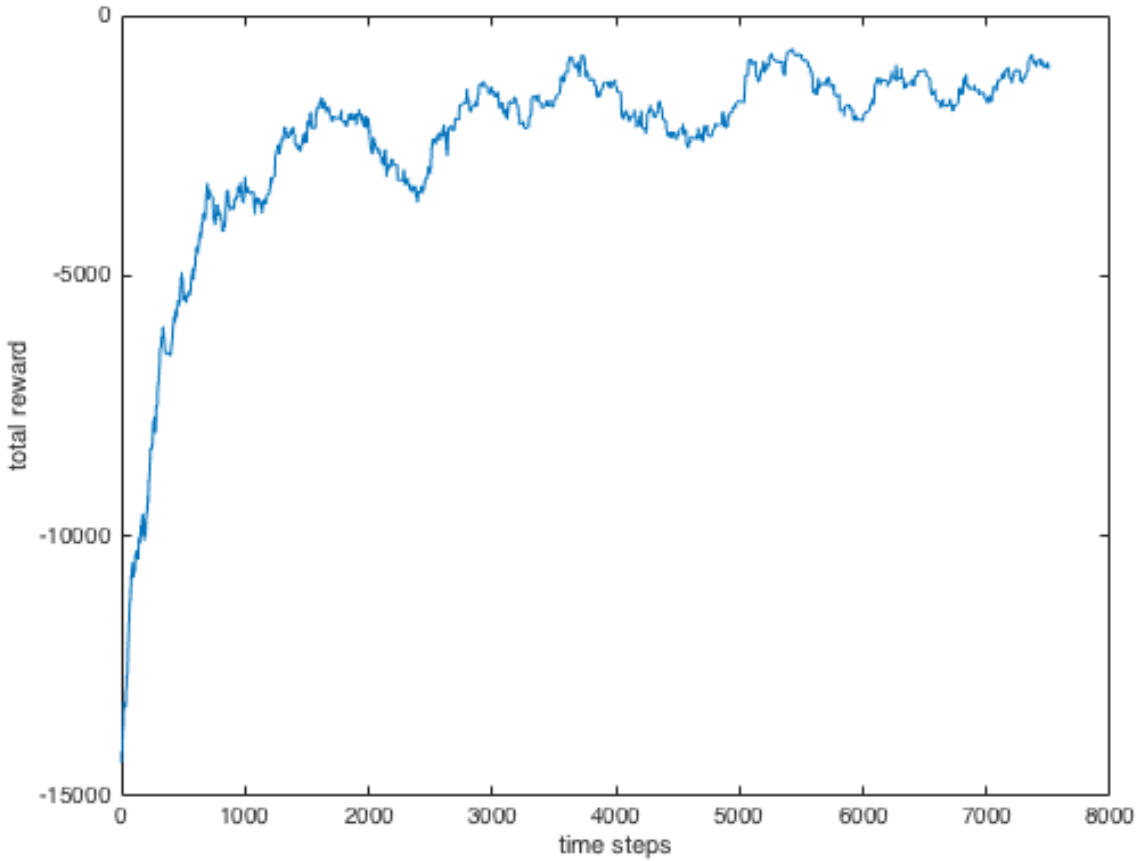
Rashej Rao (implementation)





Current Progress

Implemented reinforcement learning over grasping uncertainty


0
-5000
-10000
-15000
total reward
0
1000
2000
3000
4000
5000
6000
7000
8000
time steps