Method (continued)

10. Assign rewards to observed states $Obs(s) \rightarrow$ prior knowledge!
   - Safe environment: no update.
   - Obstacle avoidance: if $\forall s' \in Obs(s) \rightarrow Q(s', a_{\text{avoid}}) \leq 0$. 
   - Pipe following: if $\forall s' \in Obs(s) \rightarrow Q(s', a_{\text{pipe}}) \geq 0$. 
   - Combination: keep safe but do not lose pipe. 

11. Take action using LCRL and update Q-values accordingly:
   
   \[
   Q(s', a') \leftarrow Q(s', a') + \mu \left[ R(s, a) - Q(s, a) + \gamma \max_{a' \in A} Q(s', a') \right],
   \]

12. Repeat 8-11 until whole pipe inspected.

Results

1. Effect of length $l$ of equilateral triangle (case: straight unobscured pipe).

2. Localisation and navigation: test cases.

Future work

- Sophisticated localisation techniques, e.g. SLAM/Kalman filter.
- Account for energy constraints.
- Less risk-averse localisation and navigation.
- Multiagent set-up.

Acknowledgements

This project was offered by BP in the context of mini projects carried out by students of the Centre for Doctoral Training in Autonomous Intelligent Machines and Systems at the University of Oxford. We would very much like to thank the BP representatives, Mr. Joe Little, Senior Technology Innovation Manager, Mr. Richard Bailey, Senior Process Engineer and Mr. Peter Collinson, Global Environmental Response Expert, for their time and guidance.

References


Method

1. Model environment as a MDP $M = (S, s, A, P, R)$:
   - $S$ is a set of states,
   - $s \in S$ is an initial state,
   - $A$ is a set of actions,
   - $P$ is a partial probabilistic transition function,
   - $R = (R_x, R_y)$ is a reward structure.

2. Define LTL property of reachability + safety:
   \[
   \square \land \square p \land \square (p \rightarrow \Box p). \tag{1}
   \]
   "ALWAYS keep safe AND eventually find pipe AND once pipe found, ALWAYS follow pipe" 

3. Create LDBA from LTL property:

4. Create product MDP between $M$ and the automaton.

5. Initialise all voxels with same $Q(s, a)$ values.

6. Initial position $s$ of explorer: centre of equilateral triangle, on surface.

7. Explorer submerges vertically until depth $d_{\text{explorer}}$, then starts navigating.

8. Localisation using Time of Arrival (ToA) and trilateration:
   \[
   d_i = cT_i, \tag{2}
   \]
   \[
   (x - x_1)^2 + (y - y_1)^2 + z^2 = d_1^2, \tag{3}
   \]
   \[
   (x - x_2)^2 + (y - y_2)^2 + z^2 = d_2^2, \tag{4}
   \]
   \[
   (x - x_3)^2 + (y - y_3)^2 + z^2 = d_3^2. \tag{5}
   \]

9. Suggested approach
   - 1 AUV (explorer) submerges, 3 AUVs (localisers) stationary on surface in an equilateral triangle formation (GPS enabled).
   - Localisation via trilateration after exchange of acoustic links.
   - LTL to express properties.
   - Q-learning to synthesise policy.
   - Observations (sensors, camera) to assign appropriate rewards.

   This differs from classic Q-learning!

   - No luxury to learn from mistakes (expensive equipment).
   - Ability to update Q-values of not visited state (observations).
   - No need to run multiple incidences to learn - continuous learning!

Overview

An AUV submerges into deep ocean to inspect underwater infrastructure.

Underwater environment difficulties:
   - No GPS.
   - Signal attenuation.
   - Multipath fading.
   - Extreme and unpredictable currents.

Goal:
   - Keep localisation error under control.
   - Synthesise policy to safely explore environment and inspect pipeline.

Task requires: Proper decision-making and positional accuracy.

Figure 4: Obstacle avoidance. Pipe following. Combination.

Localisation very accurate in vertical, but fails in horizontal movement. Synthesised navigation policy very effective under difficult scenarios.

This project was offered by BP in the context of mini projects carried out by students of the Centre for Doctoral Training in Autonomous Intelligent Machines and Systems at the University of Oxford. We would very much like to thank the BP representatives, Mr. Joe Little, Senior Technology Innovation Manager, Mr. Richard Bailey, Senior Process Engineer and Mr. Peter Collinson, Global Environmental Response Expert, for their time and guidance.