Learning a wall following behaviour in mobile robotics using stereo and mono vision

P. Quintía J.E. Domenech C.V. Regueiro C. Gamallo[†]
R. Iglesias[†]

Dpt. Electronics and Systems. Univ. A Coruña
pquintia@udc.es, pintzio@gmail.com, cvazquez@udc.es

[†]Dpt. Electronics and Computer Science. Univ. Santiago de Compostela
cgamallo@usc.es, elberto@usc.es

IX WORKSHOP DE AGENTES FÍSICOS

Outline

Outline

1. Motivation
2. Visual behaviours
   - Reinforcement learning
   - State space definition
   - Reinforcement
3. Experimental results
   - Behaviour robustness
4. Conclusions

Motivation

- Reactive behaviours: very efficient in robotics
- **Problems**: sensors only detect obstacles on a plane
  each behaviour needs a specific adjustment

Solution

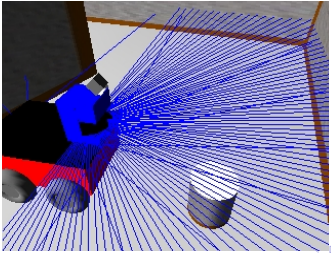
Reinforcement learning of visual behaviours

Motivation

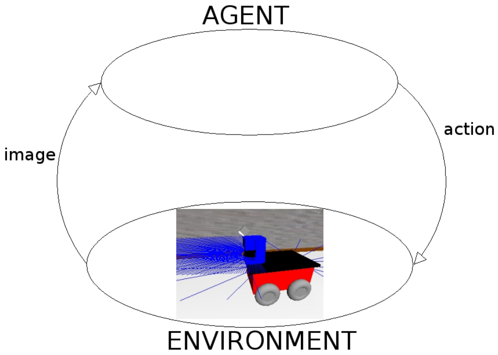
- Reactive behaviours: very efficient in robotics
- **Problems**: sensors only detect obstacles on a plane
  each behaviour needs a specific adjustment

Solution

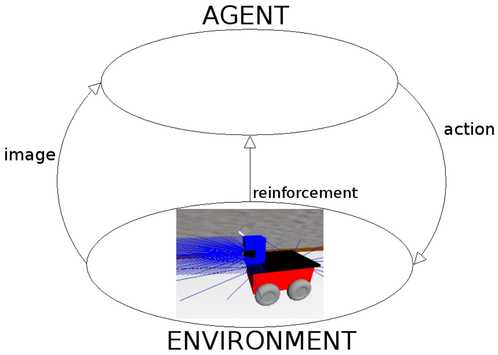
Reinforcement learning of visual behaviours



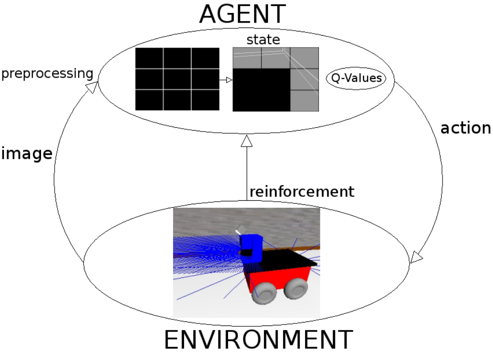
Outline



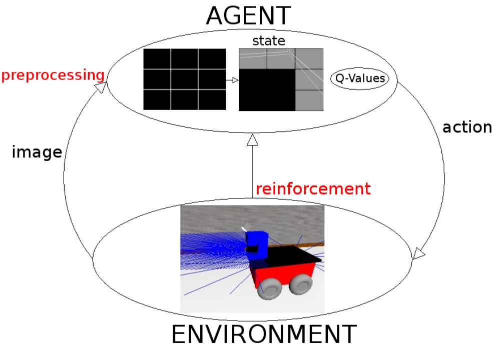
Outline



Outline


AGENT
state
preprocessing
Q-Values
action
image
reinforcement
ENVIRONMENT

Outline


AGENT
state
preprocessing
Q-Values
action
image
reinforcement
ENVIRONMENT

Related work

- Very few visual systems in mobile robotics valid for several behaviours
  - because high computational cost?
  - usually servoing and wandering
- Boada: watching, orientation and approach behaviours
  Needs to identify the objective (difficult with a contour)

Outline

Reinforcement learning (RL)

- **RL**: environment qualifies agent's actions (reward)
  - Situations are coded on states were an action is required
  - Quality function approximation $Q(s, a)$ (Q-values): best action in each state = maximum value
- **Truncatated Temporal Diferences - Q-learning**: simple and ease to implement

$$\Delta Q(s_t, a_t) = \alpha[r_t + \gamma max_a Q(s_{t+1}, a) - Q(s_t, a_t)]e_t(s)$$

Reinforcement learning (RL)

- **RL**: environment qualifies agent's actions (reward)
  - Situations are coded on states were an action is required
  - Quality function approximation $Q(s, a)$ (Q-values): best action in each state = maximum value
- **Truncatated Temporal Diferences - Q-learning**: simple and ease to implement

$$\Delta Q(s_t, a_t) = \alpha[r_t + \gamma max_a Q(s_{t+1}, a) - Q(s_t, a_t)]e_t(s)$$

State space definition

- Definition of state space is critical in RL:
  - Minimum size (converge time increase exponentially)
  - Avoid *perceptual aliasing* (unstable, no convergence)

Proposed methodology

- Regular image-based codification of states
- Behaviour independent (task-robot-environment)
- No changes between behaviours: same camera and position

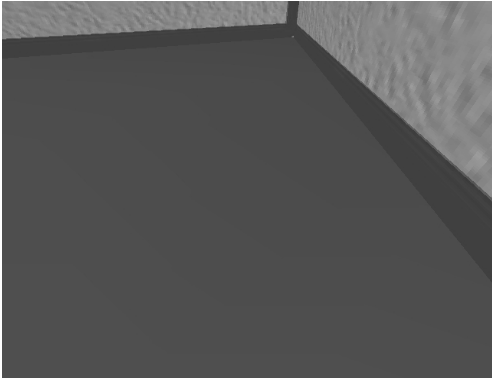
State space definition

- Definition of state space is critical in RL:
  - Minimum size (converge time increase exponentially)
  - Avoid *perceptual aliasing* (unstable, no convergence)

Proposed methodology

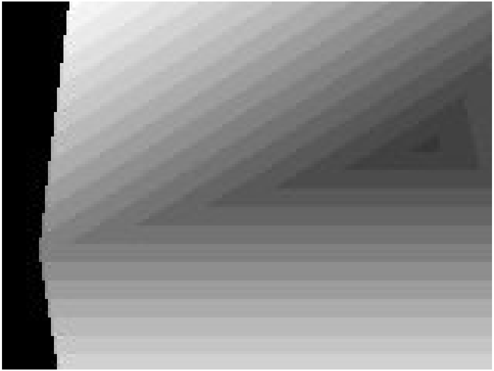
- Regular image-based codification of states
- Behaviour independent (task-robot-environment)
- No changes between behaviours: same camera and position

Image preprocessing



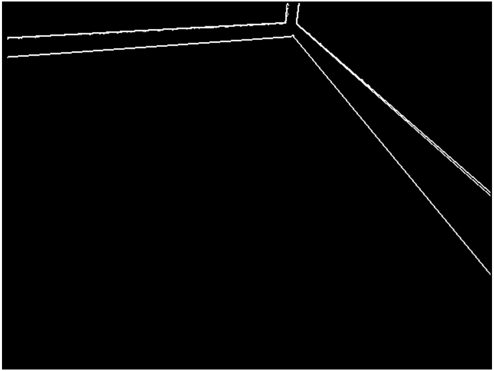
1. Filtering
   critical!!
2. Codification grid
3. Final state (free and occupied cells)

Image preprocessing



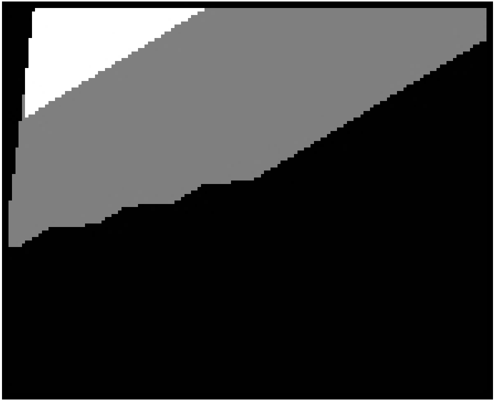
1. Filtering
   critical!!
2. Codification grid
3. Final state (free and occupied cells)

Image preprocessing



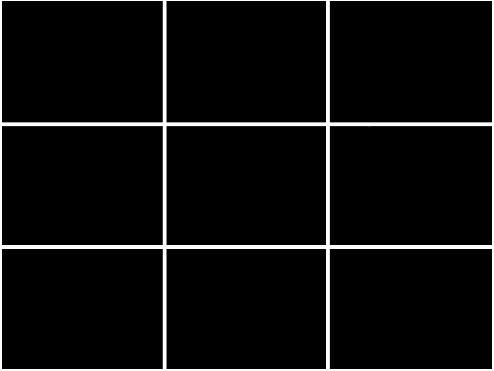
1. Filtering
   critical!!
2. Codification grid
3. Final state (free and occupied cells)

Image preprocessing



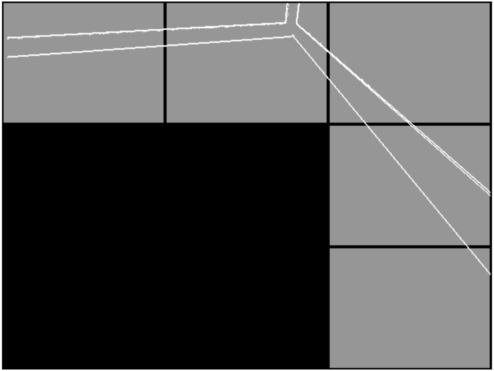
1. Filtering
   critical!!
2. Codification grid
3. Final state (free and occupied cells)

Image preprocessing



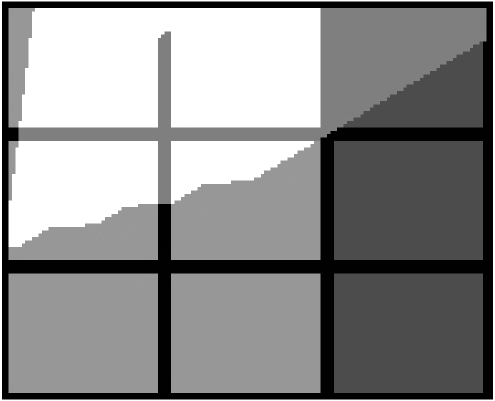
1. Filtering
   critical!!
2. Codification grid
3. Final state (free and occupied cells)

Image preprocessing



1. Filtering
   critical!!
2. Codification grid
3. Final state (free and occupied cells)

Image preprocessing



1. Filtering
   critical!!
2. Codification grid
3. Final state (free and occupied cells)

Rewards

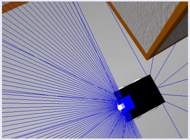
Reward

- Dependant of the behaviour wanted
- Intuitive and general definition
- Bad actions
- Negative (-1.0) and spurious

Reinforcements

Wall following

Danger of collision

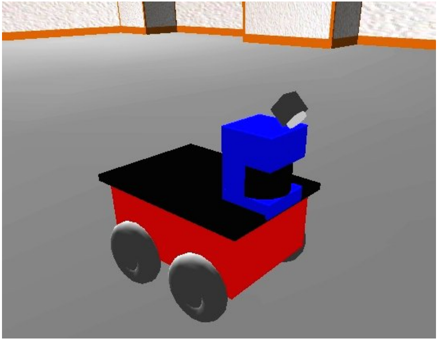


laser[0] > 1

Outline

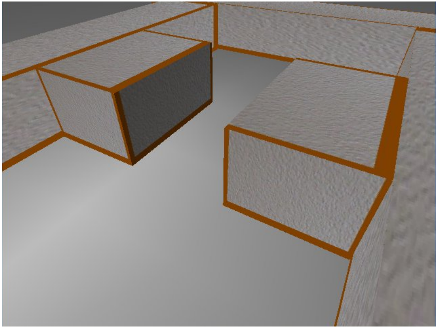
1. Motivation
2. Visual behaviours
   - Reinforcement learning
   - State space definition
   - Reinforcement
3. Experimental results
   - Behaviour robustness
4. Conclusions

Description



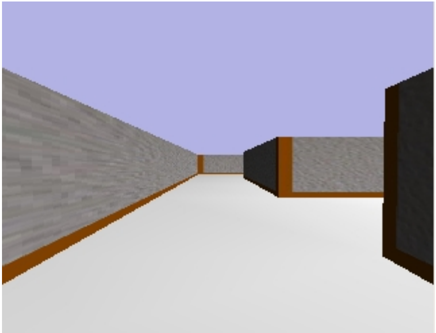
- Robot: *Pioneer 2 AT*
- 3D simulation: *Gazebo* (Player/Stage)
- Camera: FOV 85º (resolution 160x120)
- Situation:
  - 53 cm above robot
  - 40º towards the floor

Description



- Robot: *Pioneer 2 AT*
- 3D simulation: *Gazebo* (Player/Stage)
- Camera: FOV 85º (resolution 160x120)
- Situation:
  - 53 cm above robot
  - 40º towards the floor

Description



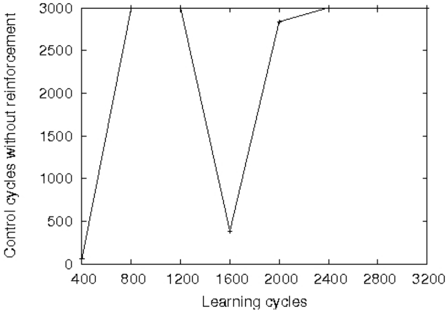
- Robot: *Pioneer 2 AT*
- 3D simulation: *Gazebo* (Player/Stage)
- Camera: FOV 85º (resolution 160x120)
- Situation:
  - 53 cm above robot
  - 40º towards the floor

Description

- Robot: *Pioneer 2 AT*
- 3D simulation: *Gazebo* (Player/Stage)
- Camera: FOV 85º (resolution 160x120)
- Situation:
  - 53 cm above robot
  - 40º towards the floor

Failure free control cycles during test

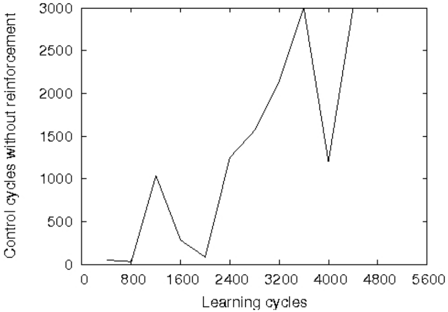
Mono vision



- 3,200 learning cycles

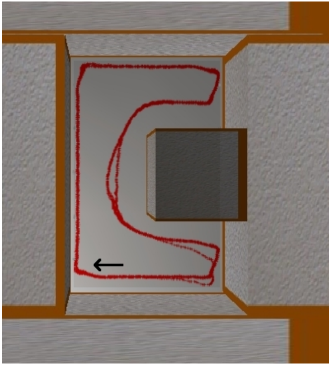
Failure free control cycles during test

Stereo vision

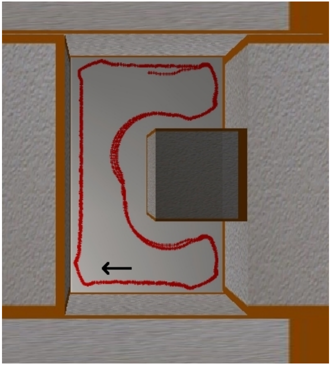


- 5,200 learning cycles

Trajectories obtained during test phase

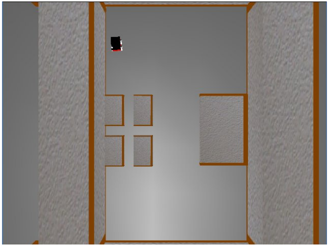


Mono vision



Stereo vision

Behaviour robustness

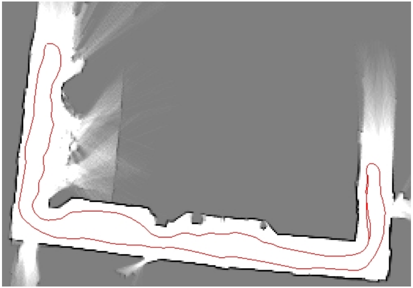


Obstacles of variable size and shape

Axis	Rotation (degrees)	Displacement (cm)
X (roll)	$[-40, 40]$	$[0, 5]$
Y (pitch)	$[0, 15]$	$[-2, 5]$
Z (yaw)	$[-3, 4]$	$[-1, 5]$

Rotations and traslations of the camera

Visual wall following behaviour with mono camera



Outline

Conclusions

- States codified directly from the image (no calibration needed) and regular 3x3 grid
- Use of TTD(λ) algorithm gives good convergence times
- Learned behaviours are robust
- Mono vision camera was tested on a real environment

Future work

- Tests with real environments and robot (stereo)
- New behaviours
- Automatic creation of state space
- Online learning

Learning a wall following behaviour in mobile robotics using stereo and mono vision

P. Quintía J.E. Domenech C.V. Regueiro C. Gamallo†
R. Iglesias†

Dpt. Electronics and Systems. Univ. A Coruña
pquintia@udc.es, pintzio@gmail.com, cvazquez@udc.es

†Dpt. Electronics and Computer Science. Univ. Santiago de Compostela
cgamallo@usc.es, elberto@usc.es

IX WORKSHOP DE AGENTES FÍSICOS