Planning to Perceive: Exploiting Mobility for Robust Object Detection

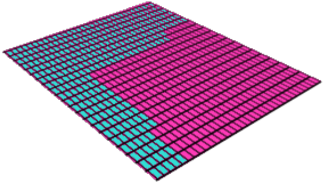
Javier Velez, Garret Hemann, Albert S. Huang, Ingmar Posner
Nicholas Roy

MIT Computer Science and Artificial Intelligence Laboratory
Mobile Robotics Group, Oxford University

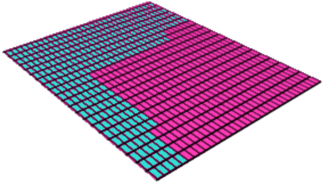
ICAPS 2011

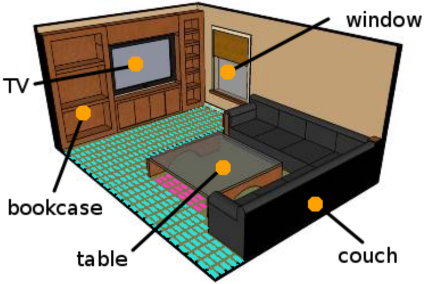
Motivation: Semantic Mapping

Motivation: Semantic Mapping



Motivation: Semantic Mapping

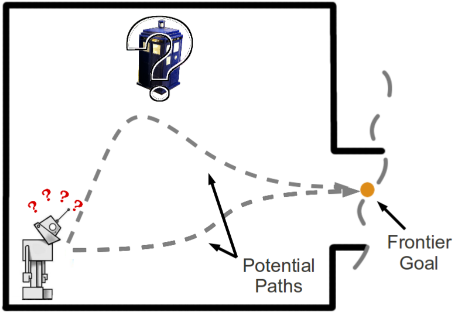




Specific Problem

Assumption

We have an exploration planner running which tells our robot new locations to explore (e.g. a frontier-based planner which tries to cover all unknown space).



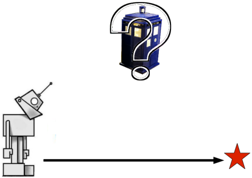
New Goal

While going to the goal location (from exploration planner), accurately detect and place objects of interest.

Exploration Approaches

Shortest Path Approach

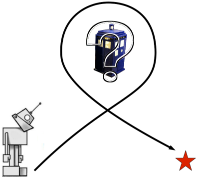
1. Go to the goal, running object detector continuously.
2. Accept object hypothesis based upon detector threshold.



Limitation: Quality of detection.

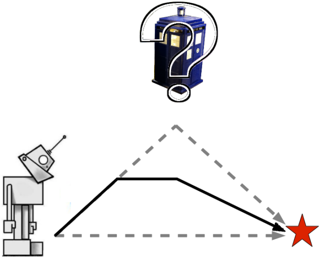
Information Gathering Approach

1. Gather as much information about the object as wanted
2. After gathering enough information, continue on towards the goal.



Limitation: Unknown path length.

Hybrid Approach



Properties

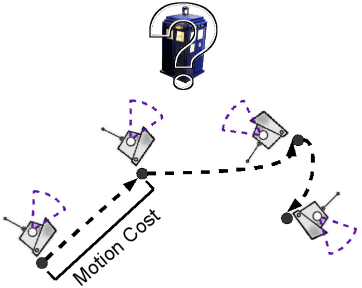
- Takes into account the motion cost of informative vantage points
- Allows for both the shortest path approach or the most information gain approach depending on cost function.
- This is a POMDP formulation [1].

The Cost of Detecting an Object

Total Cost

$$Cost(T) = C_{motion}(T) + \alpha \cdot E[C_{decision}(T)]$$

Sample Trajectory T



Decision

object exists?	*decision* accept	reject
yes	K_{tp}	K_{fp}
no	K_{fn}	K_{tn}

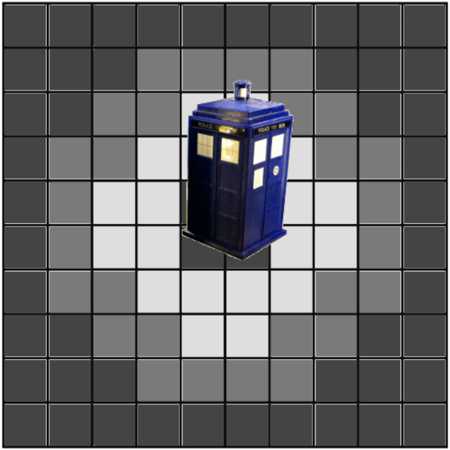
Perception Field: Expected Information Around an Object Hypothesis

Sensor Model

We build a model $p(z|o, x)$ of the detector output z given whether the object was truly there ($o = 1$) or not ($o = 0$) and the relative position x of the object.

Perception Field

Using the sensor model, we can evaluate the expected information at points around an object.



Guided Forward Search

- We can use the *perception field* to guide a forward search for good trajectories through space.
- We then pick the trajectory from those sampled with the lowest cost.
- This is a belief roadmap [3] scheme.

Initial Algorithm

Algorithm `replan_on_new_detection`

Input: an object detection z

1: update object belief with z
2: **while** planning time remains **do**
3: $T_i \sim P$ *// Sample trajectory using perception field*
4: $\mathbf{T} \leftarrow \mathbf{T} \cup T_i$
5: $T^* \leftarrow \arg\min_{T_i \in \mathbf{T}} (c_t(T_i))$
6: execute trajectory T^*

The Result: Our Robot Stands Still

Our robot stands still at the most informative location until satisfied then moves on.

The Result: Our Robot Stands Still

Our robot stands still at the most informative location until satisfied then moves on.

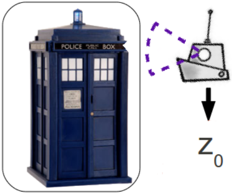
This is not what we want!

Independence

Our initial approach assumed that object detections were *independent*.

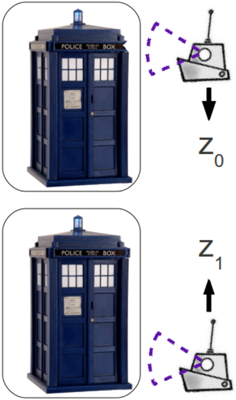
Independence

Our initial approach assumed that object detections were *independent*.



Independence

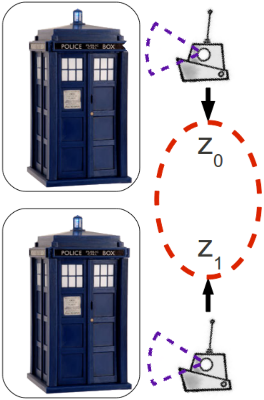
Our initial approach assumed that object detections were *independent*.



Independence

Our initial approach assumed that object detections were *independent*.

This is clearly not so!



A True Model for Object Detection

Modeling The Environment

Observations from our object detector are influenced by a global hidden variable: the environment.

A True Model for Object Detection

Modeling The Environment

Observations from our object detector are influenced by a global hidden variable: the environment.

Original

A True Model for Object Detection

Modeling The Environment

Observations from our object detector are influenced by a global hidden variable: the environment.

Original

Lighting

A True Model for Object Detection

Modeling The Environment

Observations from our object detector are influenced by a global hidden variable: the environment.

Original

Lighting

Occlusion

A True Model for Object Detection

Modeling The Environment

Observations from our object detector are influenced by a global hidden variable: the environment.

Original

Lighting

Occlusion

Disadvantages

Computational burden to model the full environment (aka. *the world*).

Planing To Perceive: Correlated Observation History

Idea

Spatial correlation between observations: locality matters most.

Properties

We no longer keep taking pictures from same vantage point since expected information diminishes because of locality correlation with previous observations.

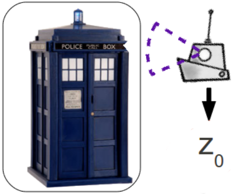
Planing To Perceive: Correlated Observation History

Idea

Spatial correlation between observations: locality matters most.

Properties

We no longer keep taking pictures from same vantage point since expected information diminishes because of locality correlation with previous observations.



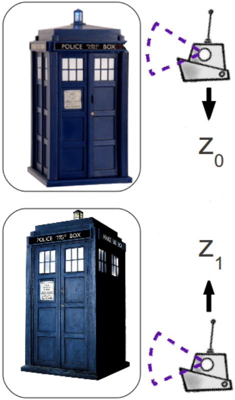
Planing To Perceive: Correlated Observation History

Idea

Spatial correlation between observations: locality matters most.

Properties

We no longer keep taking pictures from same vantage point since expected information diminishes because of locality correlation with previous observations.



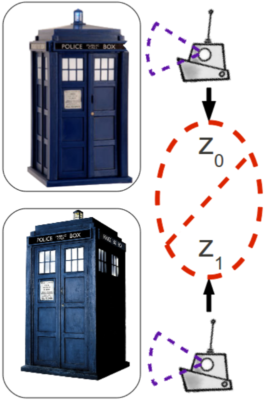
Planing To Perceive: Correlated Observation History

Idea

Spatial correlation between observations: locality matters most.

Properties

We no longer keep taking pictures from same vantage point since expected information diminishes because of locality correlation with previous observations.



Fully Correlated Input Model

- We build a model for the probability of two observations at two locations being *fully* correlated or not.

  $p(z_i = z_j | x_i, x_j)$

Fully Correlated Input Model

- We build a model for the probability of two observations at two locations being *fully* correlated or not.

  $p(z_i = z_j | x_i, x_j)$
- Observations from the same location are set to be fully correlated.

  $p(z_i = z_j | x_i = x_j) = 0$

Fully Correlated Input Model

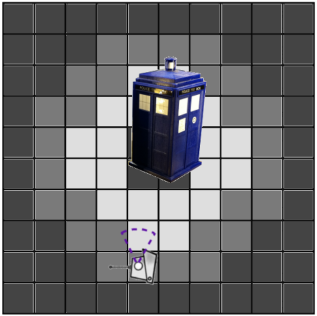
- We build a model for the probability of two observations at two locations being *fully* correlated or not.

  $p(z_i = z_j | x_i, x_j)$
- Observations from the same location are set to be fully correlated.

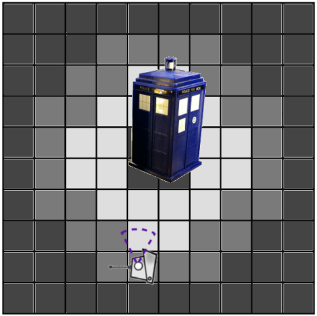
  $p(z_i = z_j | x_i = x_j) = 0$
- The probability of two locations being fully correlated is inversely proportional to the distance between the two locations.

  $$p(z_i = z_j | x_i, x_j) = \begin{cases} 1 - \frac{distance(x_1, x_j)}{d_{max}} & \Leftrightarrow d < d_{max} \\ 0 & \Leftrightarrow d \geq d_{max} \end{cases}$$

Dynamic Perception Field

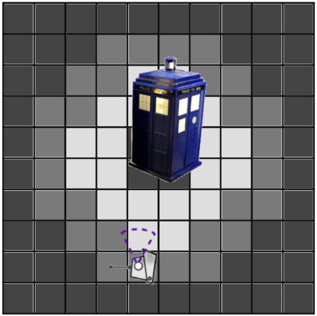


Dynamic Perception Field



$$\times \quad p(z_i = z_j | x_i, x_j)$$

Dynamic Perception Field



$$\times \quad p(z_i = z_j | x_i, x_j) \quad =$$

Planning To Perceive

Algorithm `replan_on_new_detection`

Input: an object detection z

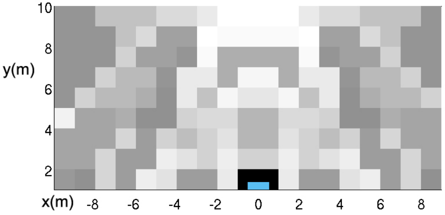
update perception field P with z
update object belief with z
while planning time remains **do**
    $T_i \sim P$ *// Sample trajectory from dynamic perception field*
    $\mathbf{T} \leftarrow \mathbf{T} \cup T_i$
$T^* \leftarrow \arg\min_{T_i \in \mathbf{T}} (c_t(T_i))$
execute trajectory T^*

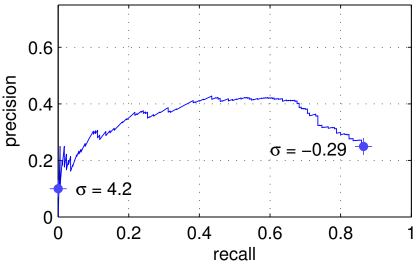
Object Detector and Perception Field

Detector

Felzenschwab [2] detector for *doors* with a stereo camera to get relative position and orientation of potential door, trained on 3500 images.

Perception Field



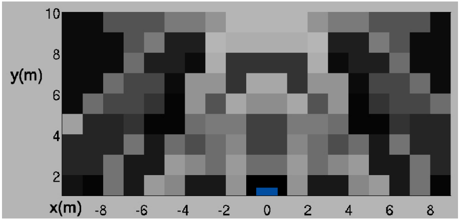


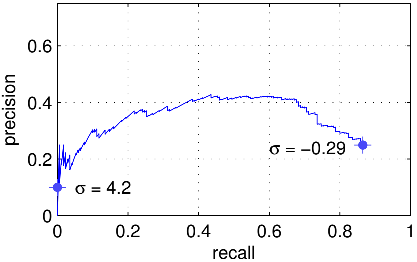
Object Detector and Perception Field

Detector

Felzenschwab [2] detector for *doors* with a stereo camera to get relative position and orientation of potential door, trained on 3500 images.

Perception Field



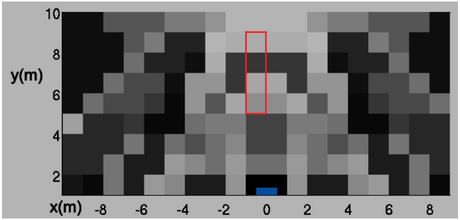


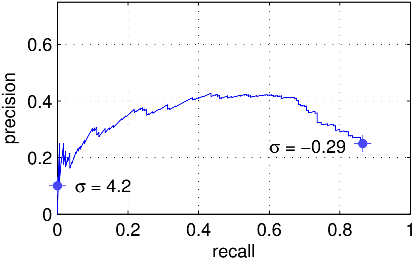
Object Detector and Perception Field

Detector

Felzenschwab [2] detector for *doors* with a stereo camera to get relative position and orientation of potential door, trained on 3500 images.

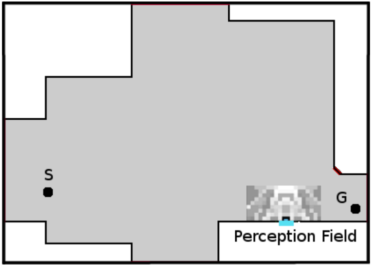
Perception Field



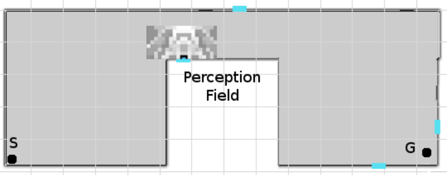


Simulation

Single Door Simulation

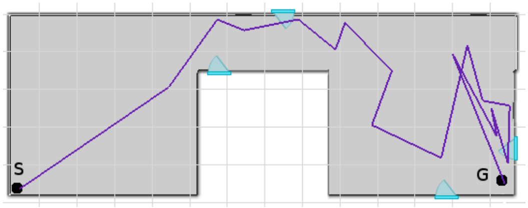


Multiple Door Simulation

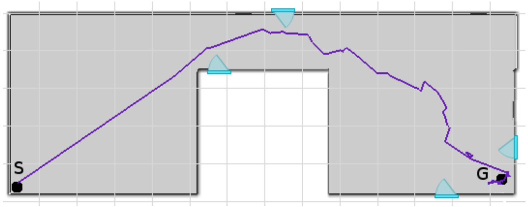


Sample Trajectories

Random Viewpoints Trajectory



Planned Viewpoints Trajectory



Results: Simulation of 50 Runs

Single Door

Average	Random
Precision	0.26 $\pm$0.04
Recall	0.60 $\pm$0.07
Path (m)	62.0 $\pm$0.67
$C_{decision}$	7.8 $\pm$0.70

Results: Simulation of 50 Runs

Single Door

Average	Random	Greedy
Precision	0.26 $\pm$0.04	0.31 $\pm$0.06
Recall	0.60 $\pm$0.07	0.44 $\pm$0.07
Path (m)	62.0 $\pm$0.67	67.0 $\pm$2.23
$C_{decision}$	7.8 $\pm$0.70	0.8 $\pm$0.83

Results: Simulation of 50 Runs

Single Door

Average	Random	Greedy	RTBSS
Precision	0.26 $\pm$0.04	0.31 $\pm$0.06	0.45 $\pm$0.06
Recall	0.60 $\pm$0.07	0.44 $\pm$0.07	0.58 $\pm$0.07
Path (m)	62.0 $\pm$0.67	67.0 $\pm$2.23	47.6 $\pm$0.19
$C_{decision}$	7.8 $\pm$0.70	0.8 $\pm$0.83	-1.6 $\pm$0.76

Results: Simulation of 50 Runs

Single Door

Average	Random	Greedy	RTBSS	Planned
Precision	0.26 $\pm$0.04	0.31 $\pm$0.06	0.45 $\pm$0.06	**0.75 $\pm$0.06**
Recall	0.60 $\pm$0.07	0.44 $\pm$0.07	0.58 $\pm$0.07	**0.80 $\pm$0.06**
Path (m)	62.0 $\pm$0.67	67.0 $\pm$2.23	**47.6 $\pm$0.19**	54.9 $\pm$3.04
$C_{decision}$	7.8 $\pm$0.70	0.8 $\pm$0.83	-1.6 $\pm$0.76	**-5.4 $\pm$0.58**

Results: Simulation of 50 Runs

Single Door

Average	Random	Greedy	RTBSS	Planned
Precision	0.26 ±0.04	0.31 ±0.06	0.45 ±0.06	**0.75 ±0.06**
Recall	0.60 ±0.07	0.44 ±0.07	0.58 ±0.07	**0.80 ±0.06**
Path (m)	62.0 ±0.67	67.0 ±2.23	**47.6 ±0.19**	54.9 ±3.04
$C_{decision}$	7.8 ±0.70	0.8 ±0.83	-1.6 ±0.76	**-5.4 ±0.58**

Results: Simulation of 50 Runs

Single Door

Average	Random	Greedy	RTBSS	Planned
Precision	0.26 $\pm$0.04	0.31 $\pm$0.06	0.45 $\pm$0.06	**0.75 $\pm$0.06**
Recall	0.60 $\pm$0.07	0.44 $\pm$0.07	0.58 $\pm$0.07	**0.80 $\pm$0.06**
Path (m)	62.0 $\pm$0.67	67.0 $\pm$2.23	**47.6 $\pm$0.19**	54.9 $\pm$3.04
$C_{decision}$	7.8 $\pm$0.70	0.8 $\pm$0.83	-1.6 $\pm$0.76	**-5.4 $\pm$0.58**

Multiple Doors

Average	Random
Precision	0.64 $\pm$0.03
Recall	0.64 $\pm$0.04
Path (m)	199 $\pm$11.24
$C_{decision}$	6.12 $\pm$1.24

Results: Simulation of 50 Runs

Single Door

Average	Random	Greedy	RTBSS	Planned
Precision	0.26 ±0.04	0.31 ±0.06	0.45 ±0.06	**0.75 ±0.06**
Recall	0.60 ±0.07	0.44 ±0.07	0.58 ±0.07	**0.80 ±0.06**
Path (m)	62.0 ±0.67	67.0 ±2.23	**47.6 ±0.19**	54.9 ±3.04
$C_{decision}$	7.8 ±0.70	0.8 ±0.83	-1.6 ±0.76	**-5.4 ±0.58**

Multiple Doors

Average	Random	Greedy
Precision	0.64 ±0.03	0.64 ±0.03
Recall	0.64 ±0.04	0.63 ±0.02
Path (m)	199 ±11.24	153 ±4.37
$C_{decision}$	6.12 ±1.24	7.32 ±1.11

Results: Simulation of 50 Runs

Single Door

Average	Random	Greedy	RTBSS	Planned
Precision	0.26 ±0.04	0.31 ±0.06	0.45 ±0.06	**0.75 ±0.06**
Recall	0.60 ±0.07	0.44 ±0.07	0.58 ±0.07	**0.80 ±0.06**
Path (m)	62.0 ±0.67	67.0 ±2.23	**47.6 ±0.19**	54.9 ±3.04
$C_{decision}$	7.8 ±0.70	0.8 ±0.83	-1.6 ±0.76	**-5.4 ±0.58**

Multiple Doors

Average	Random	Greedy	RTBSS
Precision	0.64 ±0.03	0.64 ±0.03	0.70 ±0.03
Recall	0.64 ±0.04	0.63 ±0.02	0.66 ±0.03
Path (m)	199 ±11.24	153 ±4.37	160 ±6.08
$C_{decision}$	6.12 ±1.24	7.32 ±1.11	4.64 ±6.25

Results: Simulation of 50 Runs

Single Door

Average	Random	Greedy	RTBSS	Planned
Precision	0.26 ±0.04	0.31 ±0.06	0.45 ±0.06	**0.75 ±0.06**
Recall	0.60 ±0.07	0.44 ±0.07	0.58 ±0.07	**0.80 ±0.06**
Path (m)	62.0 ±0.67	67.0 ±2.23	**47.6 ±0.19**	54.9 ±3.04
$C_{decision}$	7.8 ±0.70	0.8 ±0.83	-1.6 ±0.76	**-5.4 ±0.58**

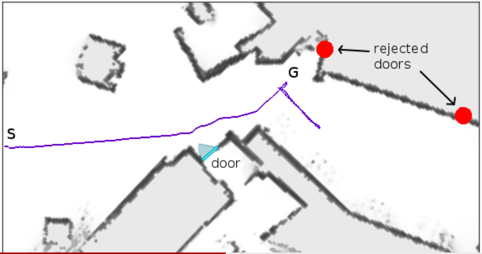
Multiple Doors

Average	Random	Greedy	RTBSS	Planned
Precision	0.64 ±0.03	0.64 ±0.03	**0.70 ±0.03**	0.53 ±0.05
Recall	0.64 ±0.04	0.63 ±0.02	0.66 ±0.03	**0.76 ±0.03**
Path (m)	199 ±11.24	153 ±4.37	160 ±6.08	**138 ±7.12**
$C_{decision}$	6.12 ±1.24	7.32 ±1.11	**4.64 ±6.25**	**4.49 ±1.37**

Results: Real World

Real World Trials

Average	Greedy	Planned
Precision	0.53 ±0.14	0.7 ±0.15
Recall	0.60 ±0.14	0.7 ±0.15
Path (m)	153.86 ±33.34	91.68 ±15.56
$C_{decision}$	1.6 ±2.61	-4.8 ±1.77
Trials	10	10



Perception Field

Conclusion

- Used cost function with takes into account both path length and information.
- Modeling correlations between object detections avoids overconfidence in expected belief update during planning.
- Our system results in higher precision and recall than a traditional object detector by itself.
- Implemented algorithm on robotic wheelchair which shows promise over other strategies to utilize an object detector

Future Work

- A correlation model that is more than just spatial correlation.
- A cost function without use of an α mixing parameter between two costs functions.

References

Frank Deinzer, Joachim Denzler, and Heinrich Niemann.
Viewpoint selection - planning optimal sequences of views for object recognition.
In *Proc. ICCV*. Springer, 2003.

P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan.
Object detection with discriminatively trained part based models.
IEEE Trans. Pattern Analysis and Machine Intelligence, 32(9), September 2010.

S. Prentice and N. Roy.
The belief roadmap: Efficient planning in belief space by factoring the covariance.
International Journal of Robotics Research, 8(11-12):1448–1465, December 2009.