IAEM Doctoral School, UL, France
LARSEN Team, LORIA, UMR 7503, France

# Robotic Coverage and Exploration Tasks as Sequential Decision-Making Problems

December 12, 2017

Presented by Nassim Kaldé

---

| | | |
|---|---|---|
| Supervisors: | Dr. François Charpillet | (LORIA/Inria Nancy) |
| | Pr. Olivier Simonin | (CITI/INSA Lyon) |
| Reviewers: | Pr. Philippe Mathieu | (LIFL/Lille 1) |
| | Pr. René Mandiau | (LAMIH/UVHC) |
| Examiners: | Dr. Laëtitia Matignon | (LIRIS/Lyon 1) |
| | Dr. Aurélie Beynier | (LIP6/UPMC) |

## Contents

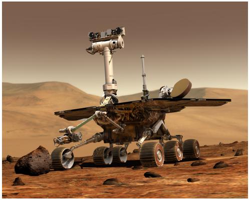
# Coverage and exploration tasks

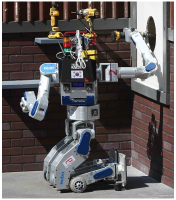
## Coverage and exploration

- Coverage: robots visit a known environment to fully observe it
- Exploration: robots visit an unknown environment to fully map it

## Recent applications (Robotics)

- NASA Mars rovers: Spirit (2003), Opportunity and Curiosity (2012)
- DARPA Robotics challenge: search and rescue, decommissioning (2012)
- iRobot Roomba vacuums: millions of robots sold (since 2002)





# Decision making in Artificial Intelligence

## Sequential Decision-Making (SDM)

- Formalisms: specify worlds with agents facing a series of choices
- Planners: optimize decisions of the agents to fulfill an objective

## Recent milestones (Game Playing)

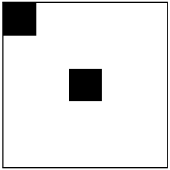
- IBM DeepBlue *vs* World Chess Champion Kasparov (1997)
- DeepMind AlphaGo *vs* World Go Champion Sedol (2016)
- CPRG DeepStack *vs* HUNL Poker pros (2017)



# Bridging coverage, exploration, and decision making

## Illustration with robot vacuums

- First time: build a map while cleaning (exploration)
- Then: clean efficiently based on the map (coverage)

## Coverage and exploration planning

- Where to go next? Information gathering
- We model coverage and exploration toy domains!

# Robotics: active sensing

## Control

| Problem | Question |
|---|---|
| Motion Planning [LaValle, 2006] | How to get from A to B? |

## State estimation

| Problem | Question |
|---|---|
| Localization [Thrun et al., 2001] | Where am I? |
| Mapping [Elfes, 1989] | Which environment is it? |
| SLAM [Durrant-Whyte and Bailey, 2006] | Where am I and in which environment? |

## Active Sensing: control and estimation [Bajcsy, 1988, Mihaylova et al., 2003]

| Problem | Where to go next to improve |
|---|---|
| Active Localization [Fox et al., 1998] | pose estimates? |
| Active Mapping [Koenig et al., 2001] | map estimates? |
| Active SLAM [Stachniss et al., 2005] | pose and map estimates? |

# Robotics: what is exploration?

## Active Mapping (AM)

How to reduce map uncertainty ...

... when localization is reliable?

## Related Work (View Candidates)

| | |
|---|---|
| Distance | [Yamauchi, 1998, Bautin et al., 2012, Faigl et al., 2012] |
| Information | [González-Banos and Latombe, 2002, Burgard et al., 2005]<br>[Amigoni and Gallo, 2005, Basilico and Amigoni, 2009]<br>[Moorehead et al., 2001, Stachniss and Burgard, 2003] |

## Drawbacks

- Good performance in general but highly suboptimal in some environments
- Greedy evaluation from current map and position
- Few theoretical guarantees [Koenig et al., 2001, Tovey and Koenig, 2003]

# Robotics: what about coverage?

## Active Coverage (AC) [Galceran and Carreras, 2013]

How to fully observe the environment when localization is reliable and ...

... measurements are known?

## Continuum of prior information

*Exploration* ⟶ *Coverage*

*Incomplete* *Environment prior* *Complete*

## However, currently...

- Active Coverage techniques fail with incomplete information
- Active Mapping techniques are blind to prior information
- Active Coverage and Mapping objectives differ

$\rightarrow$ We address both coverage and exploration in the SDM framework

# Graph Theory: what is coverage?

**Minimum Coverage Path (MCP) [Wernli, 2012, Garey and Johnson, 2002]**

- Instance: Graph $G(V,E)$, a function $w_E : E \to Z^+$, a source $s \in V$.
- Question: Find an optimal coverage path:
  A sequence $S = \langle s = v_1, v_2, \ldots, v_k \rangle$ such that $\forall i \in [1, k-1]$, $(v_i, v_{i+1}) \in E$ and $\cup_{i=1}^{k} v_i = V$, with minimum cost $\sum_{i=1}^{k-1} w_E(v_i, v_{i+1})$.
- Complexity: NP-hard

**SDM formalism [Russell and Norvig, 1995, Li et al., 2012]**

- Complete prior, deterministic control and sensing
- Completely Observable (State) Search Problem (COSP)

# Graph Theory: what about exploration?

## Canadian Traveler Problem (CTP) [Papadimitriou and Yannakakis, 1991, Fried et al., 2013]

- Instance: Graph $G(V,E)$, a function $w_E : E \rightarrow Z^+$, a source $s \in V$, a target $t \in V$, a blocking probability $p : E \rightarrow [0,1]$. Update blocking probability of neighboring edges upon node visitation.
- Question: Find an optimal navigation policy:
  A mapping $\pi : V \times |E-1|\text{-simplex} \rightarrow E$ that minimizes the expected traversal cost $C(\pi)$ from $s$ to $t$.
- Complexity: PSPACE-hard

## SDM formalism [Blei and Kaelbling, 1999, Nikolova and Karger, 2008]

- Incomplete prior, deterministic control and sensing
- Completely Observable (Belief state) Markov Decision Process ($\text{COMDP}_{baz}$)

$\rightarrow$ Exploration is a Covering CTP! [Liao and Huang, 2014]

# Graph Theory: illustration

| | Truth | Known | Used | $MCP_{\Leftarrow}$ | $MCP_{\Rightarrow}$ | Decision |
|---|---|---|---|---|---|---|
| Coverage |  |  |  | $\Rightarrow\times3$ | $\Rightarrow,\Leftarrow\times3$ | $\arg\min_{a\in\{\Leftarrow,\Rightarrow\}} Cost(MCP_a(p,m))$<br>where<br>$Cost(MCP_{\Leftarrow}(p,m))=4$<br>$Cost(MCP_{\Rightarrow}(p,m))=5$<br>hence go $\Leftarrow,\Rightarrow\times3$ |
| Exploration | |  |  | $\Rightarrow\times3$ | $\Rightarrow,\Leftarrow\times3$ | $\arg\min_{a\in\{\Leftarrow,\Rightarrow\}} \mathbb{E}_m[Cost(MCP_a(p,m)]$<br>where<br>$\mathbb{E}_m[Cost(MCP_{\Leftarrow}(p,m)]=1+16/8$<br>$\mathbb{E}_m[Cost(MCP_{\Rightarrow}(p,m)]=1+18/8$<br>hence go $\Leftarrow$ |
| | | |  | $\Rightarrow\times3$ | $\Rightarrow,\Leftarrow\times2$ | |
| | | |  | $\Rightarrow\times2$ | $\Leftarrow\times1$ | |
| | | |  | $\Rightarrow\times2$ | $\Leftarrow\times1$ | |
| | | |  | $\Rightarrow\times2$ | $\Rightarrow,\Leftarrow\times3$ | |
| | | |  | $\Rightarrow\times2$ | $\Rightarrow,\Leftarrow\times2$ | |
| | | |  | $\Rightarrow\times1$ | $\Leftarrow\times1$ | |
| | | |  | $\Rightarrow\times1$ | $\Leftarrow\times1$ | |

## Summary

| | Coverage | Exploration |
|---|---|---|
| Robotics | AC | AM |
| Graph Theory | MCP | CCTP |
| Sequential Decision-Making | COSP | $\text{COCP}_{baz}$ |

- Study 1: Modeling coverage with knowledge rewards
- Study 2: Optimal coverage with search
- Study 3: Modeling exploration with belief rewards
- Study 4: Greedy exploration with interactions

# Study 1: Models of coverage

## Objective: model coverage with SDM formalisms

1. Specify coverage domains
2. Specify coverage models

| | Truth | Known | Used | MCP$_{\Leftarrow}$ | MCP$_{\Rightarrow}$ | Decision |
|---|---|---|---|---|---|---|
| Coverage |  |  |  | $\Rightarrow\times 3$ | $\Rightarrow, \Leftarrow \times 3$ | $\arg\min_{a\in\{\Leftarrow,\Rightarrow\}} Cost(MCP_a(p,m))$<br>where<br>$Cost(MCP_{\Leftarrow}(p,m))=4$<br>$Cost(MCP_{\Rightarrow}(p,m))=5$<br>hence go $\Leftarrow, \Rightarrow \times 3$ |

## Assumptions

- Mono robot
- Static environments
- Stationary transitions
- Perfect sensing
- Complete prior

# Background on SDM formalisms

## Formalisms

- State
- Actions
- State-transition function
- State-observation function
- Rewards

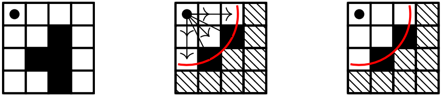
## State-transition function

- Deterministic: next state is known
- Nondeterministic: unknown probability
- Stochastic: known probability

# Coverage with deterministic control and sensing

## Deterministic control

## World state, robot field of view raycasting, deterministic sensing



## Formalism (Completely Observable (State) Search Problem (COSP))

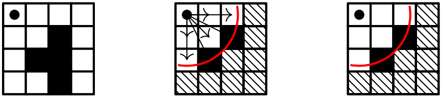
$$\mathbb{COSP} = \langle\ s_0,\ Action,\ \underline{Transition},\ Goal,\ Cost\ \rangle$$

- $s_0$, *initial state*
- $Action(s)$, *actions available from state s*
- $Transition(s,a)$, *state attained from s by action a*
- $Goal(s)$, *goal predicate for s*
- $Cost(s,a,s')$, *cost incurred from s to s′ by a*

# Coverage with deterministic control and sensing

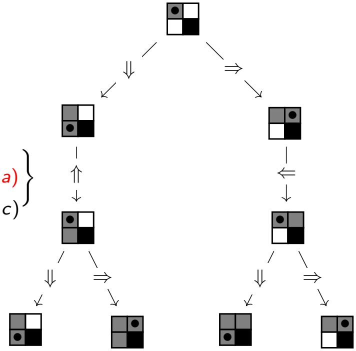
## Deterministic control

## World state, robot field of view raycasting, deterministic sensing



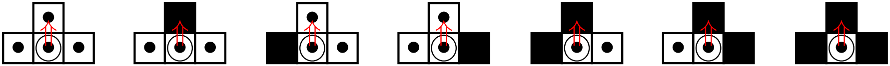
## COSP-based coverage model: $cov_{det}(map{:}\langle 2,2,m_0\rangle, robot{:}\langle p_0,0\rangle)$

- $s_0 \leftarrow \langle p_0, m_0^*, c_0\rangle$ s.t. $\begin{array}{l} p_0 \in P = [1..rows]\times[1..cols] \\ m_0 : P \to \{f\ (free), \neg f\} \\ c_0 : P \to \{c\ (covered), \neg c\} \end{array}$
- $Action(s) \subseteq \{\Uparrow, \Rightarrow, \Downarrow, \Leftarrow\}$
- $Transition(s,a) \leftarrow \left\{ s' = \langle p', m', c'\rangle \;\middle|\; \begin{array}{l} m' = nextOcc(m) \\ p' = nextPose(p, m', a) \\ c' = nextCov(p', m', c) \end{array} \right\}$
- $Goal(s) \leftarrow (s_c = c_G)$
- $Cost(s,a,s') \leftarrow \begin{cases} 0 & \text{if Goal(s)} \\ 1 - |s_c'^{-1}(c) \setminus s_c^{-1}(c)| & \text{otherwise} \end{cases}$



# Coverage with nondeterministic control and deterministic sensing

## Nondeterministic control



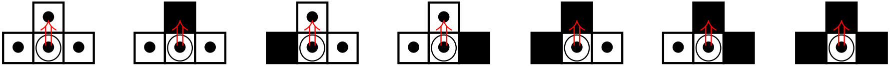
## Formalism (Completely Observable Contingency Problem (COCP))

$$\mathbb{COCP} = \langle\ s_0,\ Action,\ \underline{Transition}',\ Goal,\ Cost\ \rangle$$

- *Transition*$'(s,a)$, *states attainable from s by action a*
- *Other items as the COSP*

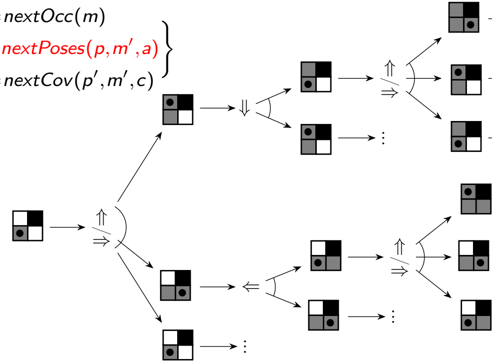
# Coverage with nondeterministic control and deterministic sensing

## Nondeterministic control



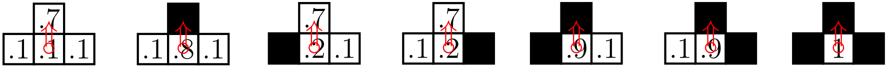
## COCP-based coverage model: $cov_{nondet}(map{:}\langle 2,2,m_0\rangle, robot{:}\langle p_0,0\rangle)$

- $Transition'(s,a) \leftarrow \left\{ s'=\langle p',m',c'\rangle \;\middle|\; \begin{array}{l} m'=nextOcc(m) \\ p'\in \color{red}{nextPoses(p,m',a)} \\ c'=nextCov(p',m',c) \end{array} \right\}$
- Other items as $cov_{det}$



# Coverage with stochastic control and deterministic sensing

## Stochastic control



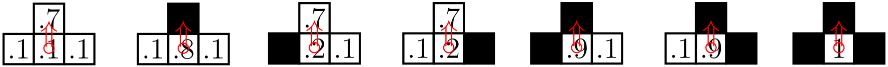

## Formalism (Completely Observable Markov Decision Process (COMDP) [Puterman, 2014])

$$\mathbb{COMDP} = \langle\ s_0,\ S,\ A,\ \underline{T},\ R\ \rangle$$

- $s_0$, *initial state*
- $S$, *finite set of states*
- $A$, *finite set of actions*
- $T(s,a,s')$, *probability of reaching $s'$ from $s$ by $a$*
- $R(s,a,s')$, *reward obtained from $s$ to $s'$ by $a$*

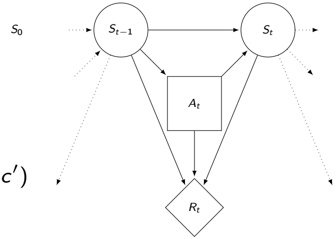
# Coverage with stochastic control and deterministic sensing

## Stochastic control



## COMDP-based coverage model: *$cov_{sto}$*

- $s_0 \leftarrow \langle p_0, m_0^*, c_0 \rangle$
- $S \leftarrow P \times M^* \times C$
- $A \leftarrow \{\Uparrow, \Rightarrow, \Downarrow, \Leftarrow\}$
- $T(s,a,s') \leftarrow \Pr(s' = \langle p', m', c' \rangle \mid s = \langle p, m, c \rangle, a)$
  $= \Pr(m' \mid m) \times \Pr(p' \mid p, m', a) \times \Pr(c' \mid p', m', c)$
  $= \mathbb{1}_{\{m\}}(m') \times \Pr(p' \mid p, m', a) \times \mathbb{1}_{\{nextCov(p', m', c)\}}(c')$
- $R(s,a,s') \leftarrow -Cost(s,a,s')$



# Conclusion

## Models of coverage

We specified three knowledge state models with deterministic sensing and

1. Deterministic control $cov_{det}$
2. Nondeterministic control $cov_{nondet}$
3. Stochastic control $cov_{sto}$

## Optimal coverage

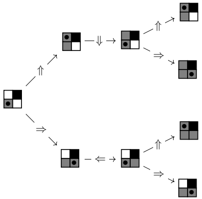
1. Complete prior
2. Nondeterministic control
3. Deterministic sensing
4. Penalize time and reward information

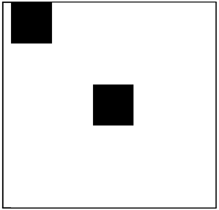
$\rightarrow$ Next, we address optimal coverage

# Study 2: Planning for Deterministic Coverage

## Objective: solving deterministic coverage with SDM planners

1. Identify standard planning techniques
2. Benchmark generic solvers and specific heuristics



## Paradigms of deterministic search

Off-line [Dijkstra, 1959]

1. Full plan
2. Full execution

On-line [Knuth, 1977]

1. Repair plan
2. Full execution

Real-time [Korf, 1990]

1. Partial plan
2. Partial execution

# Background on heuristic search

## Core principles

- Generic solvers: optimal cost-so-far computation [Bellman, 1957]

$$g(s) \approx g^*(s) = \begin{cases} 0 & \text{if } \mathit{source}(n) \\ \min_{s' \in \mathit{Pred}(s)} g^*(s') + c(s', s) & \text{otherwise} \end{cases}$$

- Specific heuristic: optimal cost-to-go estimate

$$h(s) \approx h^*(s)$$

# Algorithms and heuristics

## Generic Solvers

| Algorithm | Selection for expansion | $f-$Evaluation | $g-$Optimality |
|---|---|---|---|
| **Best-First Search Framework** | | | |
| **Iterative Best-First Search (I-BFS) [Dechter and Pearl, 1985]** | $f$-promising Open node | $agg(g,h)$ | Yes* |
| Iterative A* (I-BFS(A*)) [Hart et al., 1968, Hart et al., 1972] | – | $g+h$ | Yes* |
| Iterative Weighted A* (I-BFS(wA*)) [Pohl, 1970] | – | $g+wh$ | Yes*$_\epsilon$ |
| **Frontier Best-First Search (F-BFS) [Korf et al., 2005]** | – | $agg(g,h)$ | Yes* |
| **Recursive Best-First Search (R-BFS) [Korf, 1993]** | $F$-promising subtree root | $agg(g,h)$ | Yes* |
| **Depth-First Search Framework** | | | |
| **Depth-First Search (I-DFS,R-DFS)** | trajectory end-node successor | $agg(g,h)$ | No |
| **Iterative-Deepening Depth-First Search (ID-DFS) [Korf, 1985]** | s.t. $f(succ) \leq \Uparrow f\text{-}lowerbound$ | $agg(g,h)$ | Yes* |
| **Branch-and-Bound Depth-First Search (BnB-DFS) [Zhang and Korf, 1995]** | s.t. $f(succ) < \Downarrow f\text{-}upperbound$ | $agg(g,h)$ | Yes* |

## Specific heuristics

**Coverage: What is not yet covered?**

$$MaxCells(s) = \left\lceil \frac{|C_A^{\neg c}(s)|}{\max_{p,p' \in P} |F(p',s_m) \setminus F(p,s_m)|} \right\rceil$$

$$C_A^{\neg c}(s) = \{p \in C_A^{max} \mid s_c(p) = \neg c\}$$

**Navigation: How far is the information?**

$$MinDist(s) = \min_i \min_{v \in V_i(s)} d^*(s_p, v)$$

$$MaxDist(s) = \max_i \min_{v \in V_i(s)} d^*(s_p, v)$$

$$V_i(s) = \{p \in P_A^{max} \mid F(s_p, s_m) \cap C_A^{\neg c, i}(s) \neq \emptyset\}$$

# Experimental protocol

## Questions

1. What can we solve optimally with few resources?
2. Otherwise, can we still obtain some guarantees?

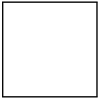
## Parameters

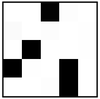
| | | |
|---|---|---|
| Instances | empty map | 291 instances |
| | not empty map | 300 instances |
| Solvers | BFS | I-BFS, R-BFS |
| | DFS | ID-DFS, BnB-DFS |
| Heuristics | coverage | MaxCells |
| | navigation | MinDist, MaxDist |
| Metrics | solution cost | path length |
| | memory consumption | node storage |
| | time consumption | node generation |
| | instance difficulty | $branching^{depth}$ |
| Resources | memory out | $1.10^6$ nodes |
| | time out | $2.10^7$ nodes |
| Details | action ordering | $\{1^{st}\Uparrow,\Rightarrow,\Downarrow,\Leftarrow\}$ |
| | second tie-breaker | lexmin $f(\cdot),h(\cdot)$ |

## Instances

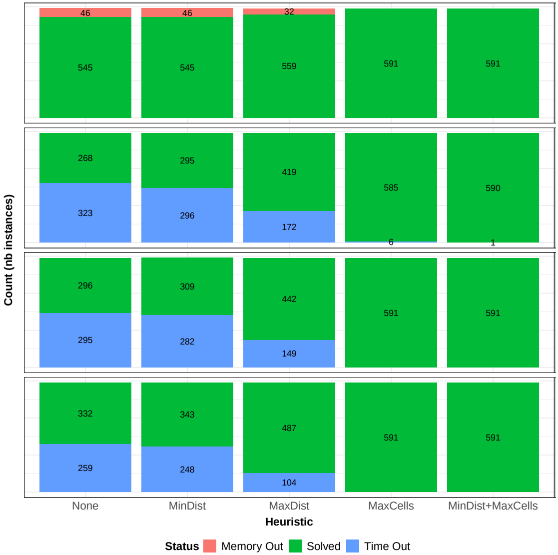
$$cov_{det}(map, robot:\langle start, range:0\rangle)$$

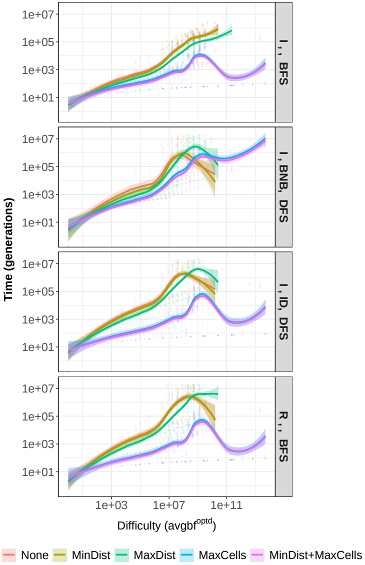
| #obstacles | #rows | #columns | #starts | #instances |
|---|---|---|---|---|
| 0 | 1 | [1-25] | [1-13] | 169 |
| 0 | 2 | [2-13] | [1-7] | 48 |
| 0 | 3 | [3-9] | [3-10] | 45 |
| 0 | 4 | [4-7] | [3-8] | 23 |
| 0 | 5 | 5 | 6 | 6 |
| 5 | 5 | 5 | 6 | 300 |





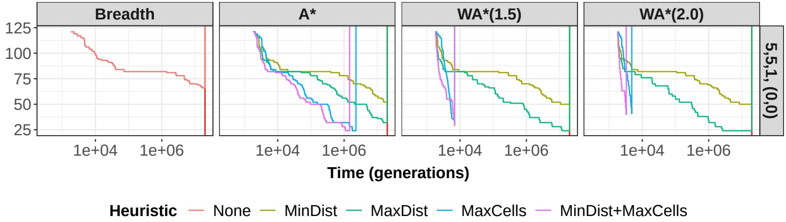
# Results: optimal solvers $f(n)=g(n)+h(n)$




*Coverage heuristics orders of magnitude faster*

# Results: anytime solver BnB-DFS(wA*)




Coverage heuristics *opt* convergence → 2-*opt* convergence ($10^2$ *faster*)

# Conclusion

## On optimal coverage $Cost(MCP)$

1. Optimal: coverage $\gg$ navigation heuristics
2. Suboptimal: guarantees $\epsilon$-optimal solutions
3. Anytime: converges to $\epsilon$-optimal coverage

## On optimal exploration $\mathbb{E}_m[Cost(MCP)]$

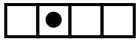
1. Coverage pattern database
2. Optimally solve exploration

$\rightarrow$ Next, we model exploration with an incomplete prior

# Study 3: Models of exploration

## Objective: Model exploration with SDM formalisms

1. Specify exploration domains
2. Specify exploration models

| | Truth | Known | Used | $MCP_{\Leftarrow}$ | $MCP_{\Rightarrow}$ | Decision |
|---|---|---|---|---|---|---|
| | | |  | $\Rightarrow \times 3$ | $\Rightarrow, \Leftarrow \times 3$ | |
| | | |  | $\Rightarrow \times 3$ | $\Rightarrow, \Leftarrow \times 2$ | $\arg\min_{a \in \{\Leftarrow, \Rightarrow\}} \mathbb{E}_m[Cost(MCP_a(p,m)]$<br>where<br>$\mathbb{E}_m[Cost(MCP_{\Leftarrow}(p,m)] = 1 + 16/8$<br>$\mathbb{E}_m[Cost(MCP_{\Rightarrow}(p,m)] = 1 + 18/8$<br>hence go $\Leftarrow$ |
| Exploration |  |  |  | $\Rightarrow \times 2$ | $\Leftarrow \times 1$ | |
| | | |  | $\Rightarrow \times 2$ | $\Leftarrow \times 1$ | |
| | | |  | $\Rightarrow \times 2$ | $\Rightarrow, \Leftarrow \times 3$ | |
| | | |  | $\Rightarrow \times 2$ | $\Rightarrow, \Leftarrow \times 2$ | |
| | | |  | $\Rightarrow \times 1$ | $\Leftarrow \times 1$ | |
| | | |  | $\Rightarrow \times 1$ | $\Leftarrow \times 1$ | |

## Assumptions

- Imperfect sensing
- Incomplete prior

# Background on SDM formalisms

## Formalisms

- State
- Actions
- State-transition function
- State-observation function
- Rewards

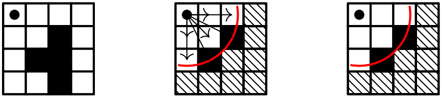
## State-observation function

- Deterministic: single measurement with certainty
- Nondeterministic: unknown probability
- Stochastic: known probability

# Domain of exploration with deterministic control and sensing

## Deterministic control

## World state, robot field of view raycasting, expected view



## Formalism (Partially Observable (State) Search Problem (POSP))

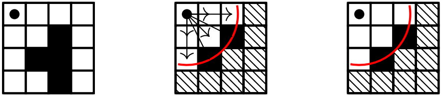
$$\mathbb{POSP} = \langle\ s_0,\ Action,\ Transition,\ Goal,\ Cost,\ \underline{Observation^{(')}}\ \rangle$$

- $\langle\ s_0,\ Action,\ Transition,\ Goal,\ Cost\ \rangle$ *is a* $\mathbb{COSP}$
- $Observation^{(')}(s,a,s')$, *potential measurements from* $s'$ *after applying* $a$ *from* $s$

# Domain of exploration with deterministic control and sensing

## Deterministic control

## World state, robot field of view raycasting, expected view

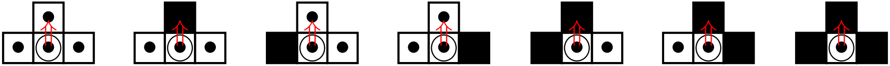


## POSP-based domain of exploration: $exp_{det}$

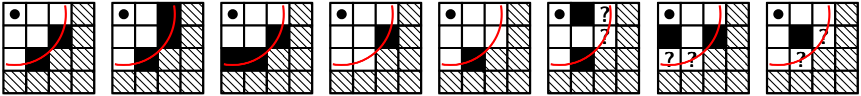
- $s_0 \leftarrow \langle p_0, m_0 \rangle$
- $Action(s) \subseteq \{\underline{noop}, \Uparrow, \Rightarrow, \Downarrow, \Leftarrow\}$
- $Transition(s,a) \leftarrow \left\{ s' = \langle p', m' \rangle \;\middle|\; \begin{array}{l} m' = nextOcc(m) \\ p' = nextPose(p, m', a) \end{array} \right\}$
- $Goal(\cdot) \leftarrow False$
- $Cost(a) \leftarrow 1 + \mathbb{1}_{\neq noop}(a)$
- $Observation(s) \leftarrow \left\{ z_s = \langle z_p, z_v \rangle \;\middle|\; \begin{array}{l} z_p = s_p \\ z_v = expectedView(s_p, s_m) \end{array} \right\}$

# Domain of exploration with nondeterministic control and sensing

## Nondeterministic control



## Nondeterministic sensing ([0..1]-error)



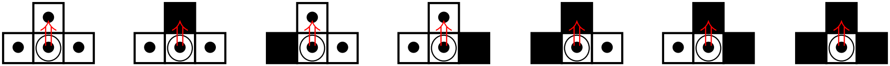

## Formalism (Partially Observable Contingency Problem (POCP))

$$\mathbb{POCP}=\langle\ s_0,\ Action,\ Transition',\ Goal,\ Cost,\ \underline{Observation^{(')}}\ \rangle$$

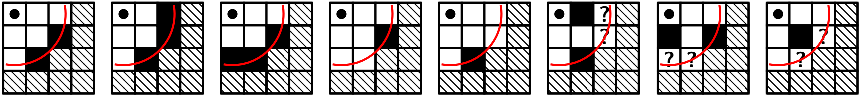
- $\langle\ s_0,\ Action,\ Transition',\ Goal,\ Cost\ \rangle$ *is a* $\mathbb{COCP}$
- $Observation^{(')}(s,a,s')$, *potential measurements from* $s'$ *after applying* $a$ *from* $s$

# Domain of exploration with nondeterministic control and sensing

## Nondeterministic control



## Nondeterministic sensing ([0..1]-error)

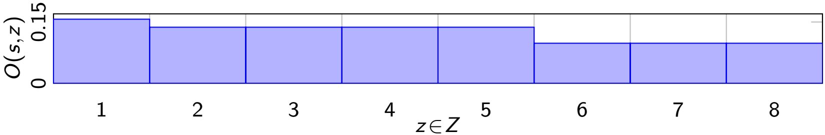



## POCP-based domain of exploration: $exp_{nondet}$

- $Transition'(s,a) \leftarrow \left\{ s' = \langle p', m' \rangle \;\middle|\; \begin{array}{l} m' = nextOcc(m) \\ p' \in nextPoses(p, m', a) \end{array} \right\}$
- $Observation'(s) \leftarrow \left\{ z_s = \langle z_p, z_v \rangle \;\middle|\; \begin{array}{l} z_p = s_p \\ z_v \in potentialViews(s_p, s_m, error_{max}) \end{array} \right\}$
- Other items as $exp_{det}$

# Domain of exploration with stochastic control and sensing

## Stochastic sensing



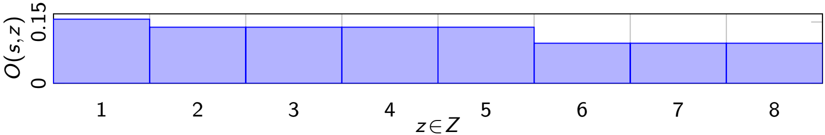

## Formalism (Partially Observable Markov Decision Process (POMDP) [Littman, 1996])

$$\mathbb{POMDP} = \langle\, s_0,\ S,\ A,\ T,\ R,\ \underline{Z},\ \underline{O}\,\rangle$$

- $\langle\, s_0, S, A, T, R\,\rangle$ *is a* $\mathbb{COMDP}$
- $Z$, *finite set of measurements*
- $O(a,s',z)$, *probability of measuring* $z$ *from state* $s'$ *after applying* $a$

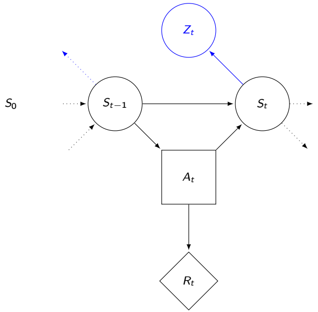
# Domain of exploration with stochastic control and sensing

## Stochastic sensing




## POMDP-based domain of exploration: $exp_{sto}$

- $s_0 : \langle p_0, m_0 \rangle$
- $S \leftarrow P \times M$
- $A \subseteq \{\underline{noop}, \Uparrow, \Rightarrow, \Downarrow, \Leftarrow\}$
- $T(s,a,s') \leftarrow \mathbb{1}_{\{m\}}(m') \times \Pr(p' \mid p, m', a)$
- $R(a) \leftarrow -Cost(a)$
- $Z \leftarrow P \times V$
- $O(s,z) \leftarrow \Pr(z = \langle p_z, v_z \rangle \mid s = \langle p, m \rangle)$
$$= \Pr(p_z \mid p) \times \Pr(v_z \mid v_{exp})$$
$$= \mathbb{1}_{\{p\}}(p_z) \times \frac{score(v, v_{exp})}{\sum_{v_z \in V(s)} score(v_z, v_{exp})}$$



# From external domains to internal models

## Domains of exploration

We specified three external state domains to simulate exploration

1. Deterministic control and sensing $exp_{det}$
2. Nondeterministic control and sensing $exp_{nondet}$
3. Stochastic control and sensing $exp_{sto}$

## Models of exploration with belief-dependent rewards

However, the agent must rely on internal belief models to handle

1. Nondeterministic sensing
2. Incomplete prior

# Model of exploration with nondeterministic control and sensing

## Formalism (Completely Observable (Belief) Contingency Problem ($COCP_{b^{az}}$))

$$COCP_{b^{az}} = \langle\ b_0,\ ApplicableAction_b,\ TransitionModel'_b,\ GoalTest_b,\ StepCost'_b\ \rangle$$

- $b_0$, *initial belief*
- $Action_b(b)$, *actions available from belief b*
- $Transition'_b(b,a)$, *beliefs attainable from b by a*
- $Goal_b(b)$, *goal predicate for b*
- $Cost'_b(b,a,z)$, *cost incurred from b after a and z*

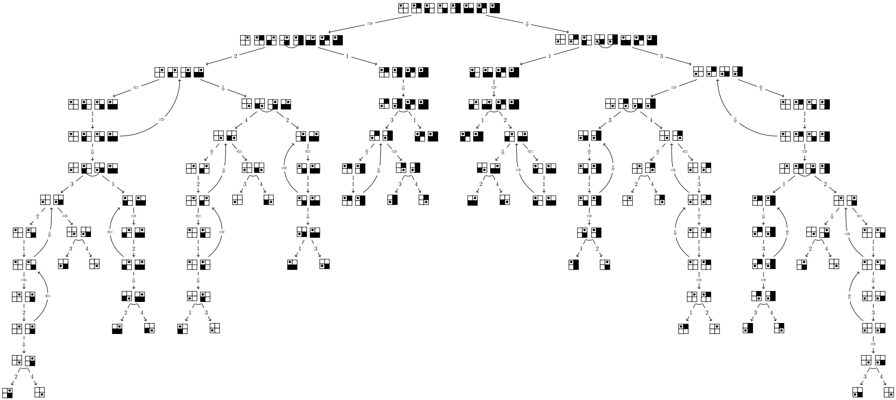
# Model of exploration with nondeterministic control and sensing

## COCP$_{b^{az}}$-based model of exploration: $exp_{(non)det}$

- $b_0 \leftarrow P_0 \times M_0$
  $P_0 = \{p_0 = \langle 1,1 \rangle\}$
  $$M_0 = \left\{ m_i \;\middle|\; m_i(p) \in \begin{cases} \{f\} & \text{if } p \in \{p_0\} \\ \{f, \neg f\} & \text{otherwise} \end{cases} \right\}$$
- $Action_b(b) \leftarrow \bigcup_{s \in b} Action(s)$
- $Goal_b \leftarrow |b| = 1$
- $Transition'_b(b,a) \leftarrow \{b^{az} \,|\, \forall z \in Z^a\}$
  $$Z^a = \bigcup_{s' \in b^a} Observation^{(')}(s')$$
  $$b^a = \bigcup_{s \in b} Transition^{(')}(s,a)$$
  $$b^{az} = \{s' \in b^a \,|\, z \in Observation^{(')}(s')\}$$
- $Cost'_b(b,a,z) \leftarrow Cost(a) - |b \setminus b^{az}|$ ($\rho$POSP/$\rho$POCP)

# Model of exploration with nondeterministic control and sensing

## COCP$_{baz}$-based belief space of exploration: $exp_{det}$



# Model of exploration with stochastic control and sensing

## Formalism (Completely Observable (Belief state) Markov Decision Process ($COMDP_{b^{az}}$))

$$COMDP_{b^{az}} = \langle\ b_0,\ B,\ A_b,\ T'_b,\ R'_b,\ \rangle$$

- $b_0$, *initial belief state*
- $B$, *continuous belief-state space*
- $A_b$, *finite set of actions*
- $T'_b(b,a,b')$, *probability of reaching* $b'$ *from* $b$ *by* $a$
- $R'_b(b,a,z)$, *reward from* $b$ *after* $a$ *and* $z$

# Model of exploration with stochastic control and sensing

## COMDP$_{b^{az}}$-based model of exploration: $exp_{sto}$

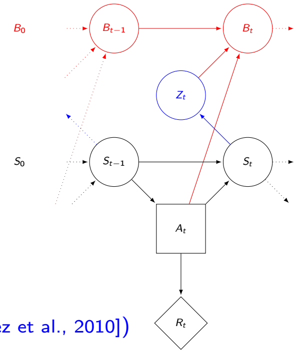
- $b_0(s) = P(S_0 = s) = \mathbf{U}_{S_0}(\{s\})$
- $B = |S-1| - simplex$
- $A_b(b) = \bigcup_{s \in S, b(s) > 0} A(s)$
- $T_b(b,a,b') = Pr(b'|b,a) = \sum_{z \in Z} \mathbb{1}_{\{b^{az}\}}(b') \Pr(z|b,a)$

$$b^a(s') = Pr(s'|b,a) = \sum_{s \in S} b(s) T(s,a,s')$$

$$b^{az}(s') = Pr(s'|b,a,z) = \eta O(s',z) b^a(s')$$

$$Pr(z|b,a) = \frac{1}{\eta} = \sum_{s' \in S} O(s',z) b^a(s')$$

- $R_b'(b,a,z) = R(a) + D_{KL}(b^{az}||b)$ ($\rho$POMDP [Araya-López et al., 2010])



# Conclusion

## Models of exploration

We reported two models with belief-dependent rewards

1. Unquantified beliefs $exp_{(non)det}$
2. Quantified beliefs $exp_{sto}$

## Optimal exploration

1. Incomplete prior
2. Nondeterministic control and sensing
3. Action penalty and information reward

## Unfortunately,...

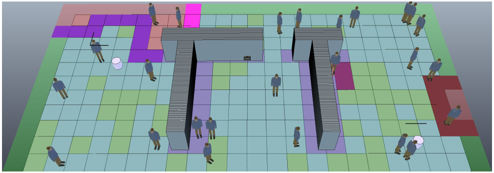
1. Pioneering algorithms do not scale [Araya-López et al., 2010]
2. Current implementation solve small instances (6 cells)
3. State-of-the-art algorithms do not reward beliefs [Silver and Veness, 2010]

→ Next, we address greedy exploration

# Study 4: Multi-Robot Exploration in Dynamic Environments

## Objective: exploring crowded indoor environments

1. Robots jointly build a map
2. Indoor crowded environments
3. Exploit pedestrian flows



## Assumptions

- Multi robot (decentralized)
- Dynamic environment
- Stationary transitions
- Perfect sensing
- Incomplete prior

# Background on greedy exploration

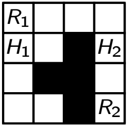
## Multi-robot exploration

- fixed:
  Hardwired [Baronov and Baillieul, 2007, Morlok and Gini, 2007]
- reactive:
  Ants [Koenig and Liu, 2001, Svennebring and Koenig, 2004, Ferranti et al., 2007, Glad et al., 2010, Andries and Charpillet, 2013]
- goal-based:
  MinDist [Yamauchi, 1997], MinPos [Bautin et al., 2012], MTSP [Faigl et al., 2012]
- utility-based:
  $V-\beta C_w$ [Burgard et al., 2005], $Ae^{-\lambda C}$ [González-Banos and Latombe, 2002], Pareto [Amigoni and Gallo, 2005], MCDM [Basilico and Amigoni, 2009], $\alpha\mathbb{E}[I]-C$ [Stachniss and Burgard, 2003]
- others:
  Auction [Zlot et al., 2002], Motivation [Macedo and Cardoso, 2004]

# Multi-robot crowded exploration domain

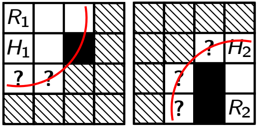
## World state

- $s\!:\!\langle m,\mathbf{R},\mathbf{H}\rangle$ s.t. $m\!:\!P\!\rightarrow\!\{f,\neg f_s\}$, map
  $\mathbf{R}\!=\!\{R_1,..,R_n\}$, robots
  $\mathbf{H}\!=\!\{H_1,..,H_m\}$, pedestrians



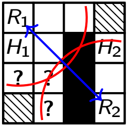

## Robot measurement

- $Z_i\!:\!\langle v_Z,R_Z,H_Z\rangle$ s.t. $v_Z\!:\!FOV_i\!\rightarrow\!\{f,\neg f_s,\neg f_m,\neg k\}$
  $R_Z\!\subseteq\!\mathbf{R}$, visible robots
  $H_Z\!\subseteq\!\mathbf{H}$, visible pedestrians




## Robot communication

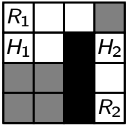
- $\Theta_i^{0:t}\!=\!\bigcup_{j=1}^{|\mathbf{R}|} Z_j^{0:t_j}$ s.t. $Z_i^{0:t}$, robot history



# Multi-robot crowded exploration model

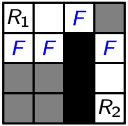
## Robot knowledge map

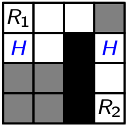
- $k_i : P \rightarrow \{f, \neg f_s, \neg f_m, \neg k\}$




## Targets $T$

- Frontiers $F$ and Interactions $H$




## On-line Task-Allocation Framework

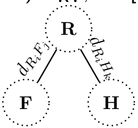
| $c_{R_iT_j}$ | $\mathbf{T}$ |
|---|---|
| $\mathbf{R}$ | $\mathbf{C_{RT}}$ |

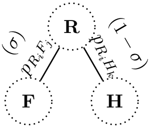
$\xrightarrow{\text{opt.}}$

| $a_{R_iT_j}$ | $\mathbf{T}$ |
|---|---|
| $\mathbf{R}$ | $\mathbf{A_{RT}}$ |

## Robot-Target assignment costs $\mathbf{C_{RG}}$

- $\mathbf{C_{RT}} = \alpha \cdot \mathbf{D_{RT}} + (1-\alpha) \cdot \mathbf{P_{RT}},\ \alpha \in [0,1]$



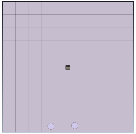
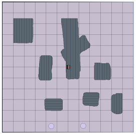
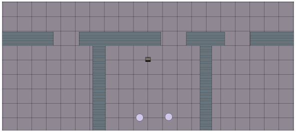

$d_{R_iX_j} = shortest\_path\_length_{R_i,X_j}$ $\quad p_{R_iX_j} = idle\_time_{R_iX_j} + reorientation_{R_iX_j}$

# Experimental protocol

## Parameters

| | | |
|---|---|---|
| Maps | empty | $10\times10$ cells ($25m^2$) |
| | unstructured | $12\times12$ cells ($36m^2$) |
| | structured | $22\times11$ cells ($60.5m^2$) |
| Strategies | greedy | local, group |
| Pedestrian | speed | $0.56m/s$ |
| | controller | V-REP Walking Bill |
| | starting | uniform random |
| Robot | speed | $0.5m/s$ |
| | viewing range | $2m$ |
| | network range | $100m$ |
| | starting | fixed positions |
| Modulators | $\alpha$ | 0:0.25:1 |
| | $\sigma$ | 0:0.25:1 |
| Metrics | coverage | $\frac{covered\ \ cells}{all\ \ cells}$ |
| | distance | $m$ |
| | time | $s$ |
| Resources | time out | $300s$ |

## Maps

## Greedy strategies

**Local: Only me**

1. Initialize $\mathbf{T}\leftarrow\{\mathbf{H}\cup\mathbf{F}\}$
2. Find $T^*\leftarrow\arg\min_{T_j\in\mathbf{T}} c_{R_i T_j}$

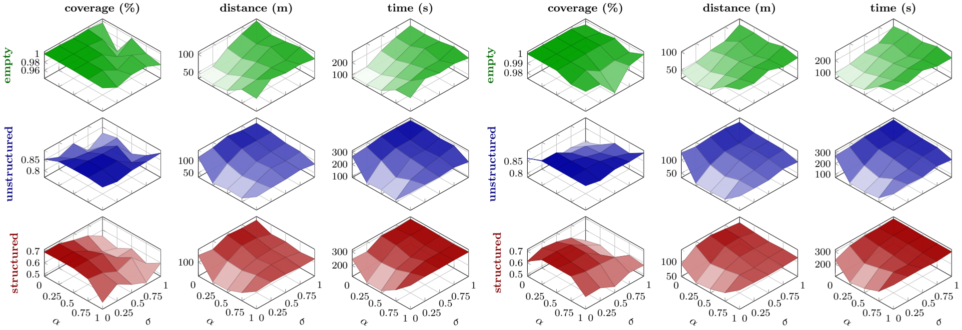
**Group: And the others?**

1. Initialize $\mathbf{T}\leftarrow\{\mathbf{H}\cup\mathbf{F}\}$; $R_V\leftarrow R_Z$
2. Find $(R^*, T^*)\leftarrow\arg\min_{R_j\in R_V, T\in\mathbf{T}} c_{R_j T_j}$
3. If $(R^*=R_i)$ then return $T^*$; else $R_Z\leftarrow R_Z\setminus R^*$; $\mathbf{T}\leftarrow\mathbf{T}\setminus T^*$
4. Go to step 2 unless $\mathbf{T}$ is empty

# Results: density (30%), robot average (20 samples)

Local Greedy

Group Greedy




Best parameters (0,0);(0.75,0);(0.5,0)

Best parameters (0.25,0);(0.5,0);(0.25,0)

frontiers are never penalized but pedestrians are

# Conclusion

## Hybrid Exploration

1. Enabled the frontier paradigm in crowded environment
2. Proposed an interactive paradigm to exploit affordances

## Early Results

1. Targets: global frontiers and local interactions
2. Costs: weighted distance and time-orientation penalty
3. Static tuning: best results when interactions are penalized

$\rightarrow$ Finally, we conclude

# Summary

## Studies 1 and 3: Modeling coverage and exploration with SDM formalisms

- Lookahead models with knowledge/belief rewards
- From complete to incomplete prior
- From deterministic to stochastic control and observation
- Allow **optimal** centralized coverage and exploration

## Study 2: Optimally solving coverage with SDM planners

- Benchmark of generic solvers and expert heuristics
- Orders of magnitude domination of coverage heuristics
- Anytime improvement of exploration trajectories

## Study 4: Greedy exploration with interactions

- Crowded environments
- Hybrid exploration paradigm
- Frontiers preferred over interactions

# Perspectives

## Studies 1 and 3: Modeling coverage and exploration with SDM formalisms

- Centralized multi-robot: $\rho$POMDP [Araya-López et al., 2010]
- Decentralized multi-robot: Dec-$\rho$POMDP [Renoux, 2015]

## Study 2: Optimally solving coverage with SDM planners

- Evaluate other heuristics: MTSP [Faigl et al., 2012]
- Consider other paradigms: On-Line [Knuth, 1977], Real-Time search [Korf, 1990]

## Study 4: Greedy exploration with interactions

- Realistic pedestrian behavior: Social Force Model [Helbing and Molnar, 1995]
- On-line parameter tuning: Embodied Evolution [Watson et al., 2002]

# The End

Thank you for your attention. Questions?

List of contributions

1. Models of coverage and exploration
2. Planning for optimal coverage
3. Greedy exploration in crowds

# References I


Amigoni, F. and Gallo, A. (2005).
A Multi-Objective Exploration Strategy for Mobile Robots.
In *Proceedings of the International Conference on Robotics and Automation*.

Andries, M. and Charpillet, F. (2013).
Multi-Robot Exploration of Unknown Environments with Identification of Exploration Completion and Post-Exploration Rendez-Vous using Ant Algorithms.
In *Proceedings of the International Conference on Intelligent Robots and Systems*.

Araya-López, M., Buffet, O., Thomas, V., and Charpillet, F. (2010).
A POMDP Extension with Belief-Dependent Rewards.
In *Advances in Neural Information Processing Systems*.

Bajcsy, R. (1988).
Active Perception.
*Proceedings of the IEEE*, 76(8).

# References II


Baronov, D. and Baillieul, J. (2007).
Reactive Exploration through Following Isolines in a Potential Field.
In *Proceedings of the American Control Conference*.

Basilico, N. and Amigoni, F. (2009).
Exploration Strategies based on Multi-Criteria Decision Making for an Autonomous Mobile Robot.
In *Proceedings of the European Conference on Mobile Robots*, pages 259–264.

Bautin, A., Simonin, O., and Charpillet, F. (2012).
MinPos : A Novel Frontier Allocation Algorithm for Multi-robot Exploration.
In *Proceedings of the International Conference on Intelligent Robotics and Applications*.

Bellman, R. (1957).
*Dynamic Programming*.
Princeton University Press.

# References III


Blei, D. M. and Kaelbling, L. P. (1999).
Shortest Paths in a Dynamic Uncertain Domain.
In *IJCAI Workshop on Adaptive Spatial Representations of Dynamic Environments*, volume 4, page 2.

Burgard, W., Moors, M., Stachniss, C., and Schneider, F. (2005).
Coordinated Multi-Robot Exploration.
*Transactions on Robotics*, 21(3).

Dechter, R. and Pearl, J. (1985).
Generalized Best-First Search Strategies and the Optimality of A*.
*Journal of the ACM*, 32(3).

Dijkstra, E. (1959).
A Note on Two Problems in Connexion with Graphs.
*Numerische mathematik*, 1(1).

Durrant-Whyte, H. and Bailey, T. (2006).
Simultaneous Localization and Mapping: Part I.
*Robotics & Automation Magazine, IEEE*, 13(2).

# References IV


Elfes, A. (1989).
Using Occupancy Grids for Mobile Robot Perception and Navigation.
*Computer*, 22(6).

Faigl, J., Kulich, M., and Přeučil, L. (2012).
Goal Assignment using Distance Cost in Multi-Robot Exploration.
In *Proceedings of the International Conference on Intelligent Robots and Systems*.

Ferranti, E., Trigoni, N., and Levene, M. (2007).
Brick&Mortar: an On-Line Multi-Agent Exploration Algorithm.
In *Proceedings of the International Conference on Robotics and Automation*.

Fox, D., Burgard, W., and Thrun, S. (1998).
Active Markov Localization for Mobile Robots.
*Robotics and Autonomous Systems*, 25(3-4).

# References V


Fried, D., Shimony, S. E., Benbassat, A., and Wenner, C. (2013).
Complexity of Canadian Traveler Problem Variants.
*Theoretical Computer Science*, 487:1–16.

Galceran, E. and Carreras, M. (2013).
A Survey on Coverage Path Planning for Robotics.
*Robotics and Autonomous Systems*, 61(12).

Garey, M. and Johnson, D. (2002).
*Computers and Intractability*, volume 29.

Glad, A., Simonin, O., Buffet, O., and Charpillet, F. (2010).
Influence of Different Execution Models on Patrolling Ant Behaviors: from Agents to Robots.
In *International Conference on Autonomous Agents and Multiagent Systems*.

González-Banos, H. H. and Latombe, J.-C. (2002).
Navigation Strategies for Exploring Indoor Environments.
*The International Journal of Robotics Research*, 21(10-11).

# References VI


Hart, P., Nilsson, N., and Raphael, B. (1968).
A Formal Basis for the Heuristic Determination of Minimum Cost Paths.
*Transactions on Systems Science and Cybernetics*, 4(2).

Hart, P., Nilsson, N., and Raphael, B. (1972).
Correction to a Formal Basis for the Heuristic Determination of Minimum Cost Paths.
*ACM SIGART Bulletin*, (37).

Helbing, D. and Molnar, P. (1995).
Social Force Model for Pedestrian Dynamics.
*Physical review E*, 51(5).

Knuth, D. (1977).
A Generalization of Dijkstra's Algorithm.
*Information Processing Letters*, 6(1).

Koenig, S. and Liu, Y. (2001).
Terrain Coverage with Ant Robots: A Simulation Study.
In *Proceedings of the international conference on Autonomous agents*.

# References VII


Koenig, S., Tovey, C., and Halliburton, W. (2001).
Greedy Mapping of Terrain.
In *Proceedings of the International Conference on Robotics and Automation*, volume 4.

Korf, R. (1985).
Depth-First Iterative-Deepening: An Optimal Admissible Tree Search.
*Artificial intelligence*, 27(1).

Korf, R. (1990).
Real-Time Heuristic Search.
*Artificial intelligence*, 42(2).

Korf, R. (1993).
Linear-Space Best-First Search.
*Artificial Intelligence*, 62(1).

Korf, R., Zhang, W., Thayer, I., and Hohwald, H. (2005).
Frontier Search.
*Journal of the ACM*, 52(5).

# References VIII


LaValle, S. M. (2006).
*Planning Algorithms*.
Cambridge university press.

Li, A. Q., Amigoni, F., and Basilico, N. (2012).
Searching for Optimal Off-Line Exploration Paths in Grid Environments for a Robot with Limited Visibility.
In *AAAI*.

Liao, C.-S. and Huang, Y. (2014).
The Covering Canadian Traveller Problem.
*Theoretical Computer Science*, 530(Supplement C):80 – 88.

Littman, M. (1996).
*Algorithms for Sequential Decision Making*.
PhD thesis, Brown University.

# References IX


- Macedo, L. and Cardoso, A. (2004).
  Exploration of Unknown Environments with Motivational Agents.
  In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems*.

- Mihaylova, L., Lefebvre, T., Bruyninckx, H., Gadeyne, K., and De Schutter, J. (2003).
  A Comparison of Decision Making Criteria and Optimization Methods for Active Robotic Sensing.
  In *Numerical Methods and Applications*.

- Moorehead, S. J., Simmons, R., and Whittaker, W. L. (2001).
  Autonomous Exploration using Multiple Sources of Information.
  In *Proceedings of the International Conference on Robotics and Automation*, volume 3.

- Morlok, R. and Gini, M. (2007).
  Dispersing Robots in an Unknown Environment.
  In *Distributed Autonomous Robotic Systems*.

# References X


Nikolova, E. and Karger, D. (2008).
Route Planning under Uncertainty: The Canadian Traveller Problem.
In *AAAI*.

Papadimitriou, C. and Yannakakis, M. (1991).
Shortest Paths without a Map.
*Theoretical Computer Science*, 84(1).

Pohl, I. (1970).
Heuristic search viewed as path finding in a graph.
*Artificial Intelligence*, 1(3).

Puterman, M. (2014).
*Markov Decision Processes: Discrete Stochastic Dynamic Programming*.

Renoux, J. (2015).
*Contribution to Multiagent Planning for Active Information Gathering*.
PhD thesis.

Russell, S. and Norvig, P. (1995).
Artificial Intelligence: Modern Approach.

Silver, D. and Veness, J. (2010).
Monte-Carlo Planning in Large POMDPs.
In *Advances in neural information processing systems*.

Stachniss, C. and Burgard, W. (2003).
Exploring Unknown Environments with Mobile Robots using Coverage Maps.
In *Proceedings of the International Joint Conference on Artificial Intelligence*.

Stachniss, C., Grisetti, G., and Burgard, W. (2005).
Information Gain-based Exploration Using Rao-Blackwellized Particle Filters.
In *Robotics: Science and Systems*, volume 2.

Svennebring, J. and Koenig, S. (2004).
Building Terrain-Covering Ant Robots: A Feasibility Study.
*Autonomous Robots*, 16(3).

# References XII


Thrun, S., Fox, D., Burgard, W., and Dellaert, F. (2001).
Robust Monte Carlo Localization for Mobile Robots.
*Artificial Intelligence*, 128(1-2).

Tovey, C. and Koenig, S. (2003).
Improved Analysis of Greedy Mapping.
In *Proceedings of the International Conference on Intelligent Robots and Systems*, volume 4.

Watson, R. A., Ficici, S. G., and Pollack, J. B. (2002).
Embodied Evolution: Distributing an Evolutionary Algorithm in a Population of Robots.
*Robotics and Autonomous Systems*, 39(1).

Wernli, D. (2012).
Grid Exploration.
Master's thesis, ETH Zürich, Department of Computer Science.

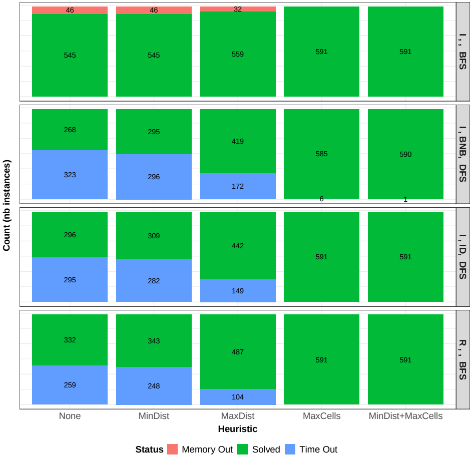
Yamauchi, B. (1997).
A Frontier-Based Approach for Autonomous Exploration.
In *Proceedings of the International Symposium on Computational Intelligence in Robotics and Automation*.

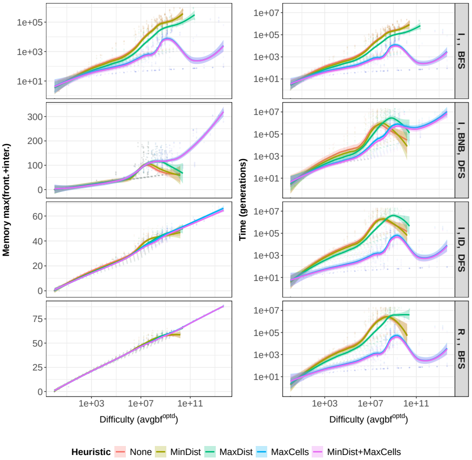
Yamauchi, B. (1998).
Frontier-based Exploration using Multiple Robots.
In *Proceedings of the International Conference on Autonomous agents*.

Zhang, W. and Korf, R. (1995).
Performance of Linear-Space Search Algorithms.
*Artificial Intelligence*, 79(2).

Zlot, R., Stentz, A., Dias, M. B., and Thayer, S. (2002).
Multi-robot Exploration Controlled by a Market Economy.
In *Proceedings of the International Conference on Robotics and Automation*, volume 3.

# Results: optimal solvers $f(n)=g(n)+h(n)$ 1/2

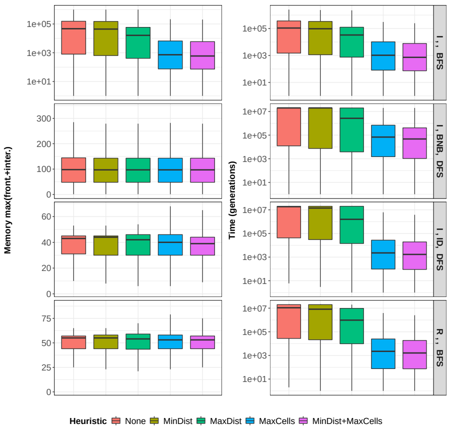


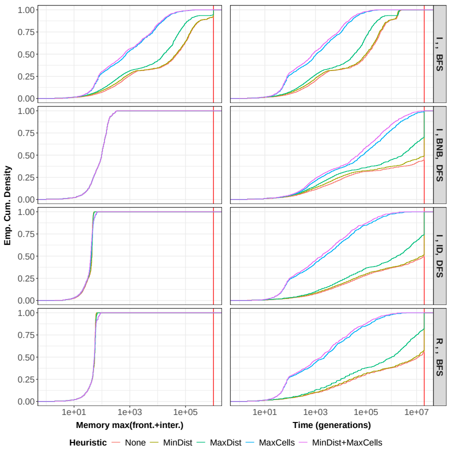




*MinDist < MaxDist < MaxCells < MinDist + MaxCells*

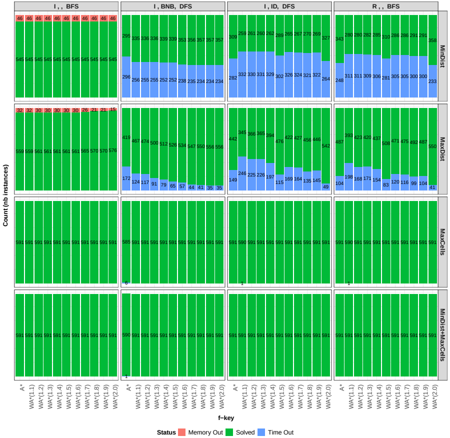
*Orders of magnitude faster Coverage ≫ Navigation*

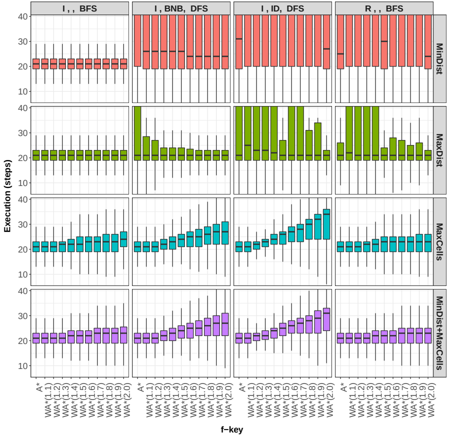
# Results: optimal solvers $f(n)=g(n)+h(n)$ 2/2



# Results: suboptimal solvers $f(n)=g(n)+wh(n)$ 1/2

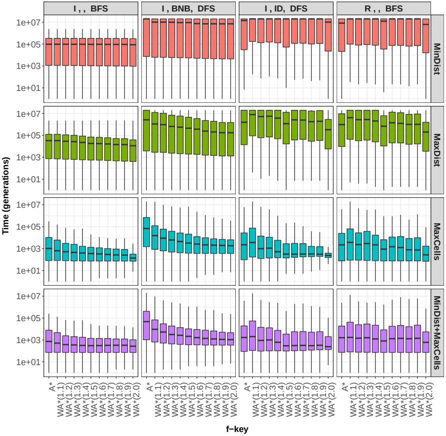


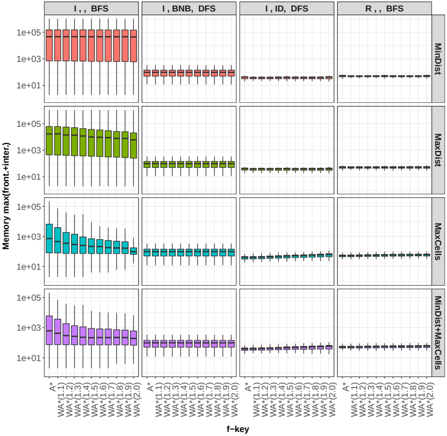




Monotonic I-BFS BnB-DFS, Nonmonotonic ID-DFS, R-BFS

1-opt in theory → better in practice

# Results: suboptimal solvers $f(n)=g(n)+wh(n)$ 2/2



# Results: anytime solver BnB-DFS(wA*)

*Opt* convergence→2-*Opt* convergence ($10^2$ *faster*)