Active Sensing as Bayes-Optimal Sequential Decision-Making

Sheeraz Ahmad & Angela J. Yu

Department of Computer Science and Engineering
University of California, San Diego

December 7, 2012

Outline

- Introduction
- Active Sensing: Background
- Visual Search Task
- POMDP Formulation
  - Bayesian Inference
  - Optimal Action Selection
  - Results
  - Scalability Issues
- Low Dimensional Approximate Control
  - Results
- Comparison with Infomax Policy
- Conclusion

Outline

Introduction

Active Sensing: Background

Visual Search Task

POMDP Formulation
- Bayesian Inference
- Optimal Action Selection
- Results
- Scalability Issues

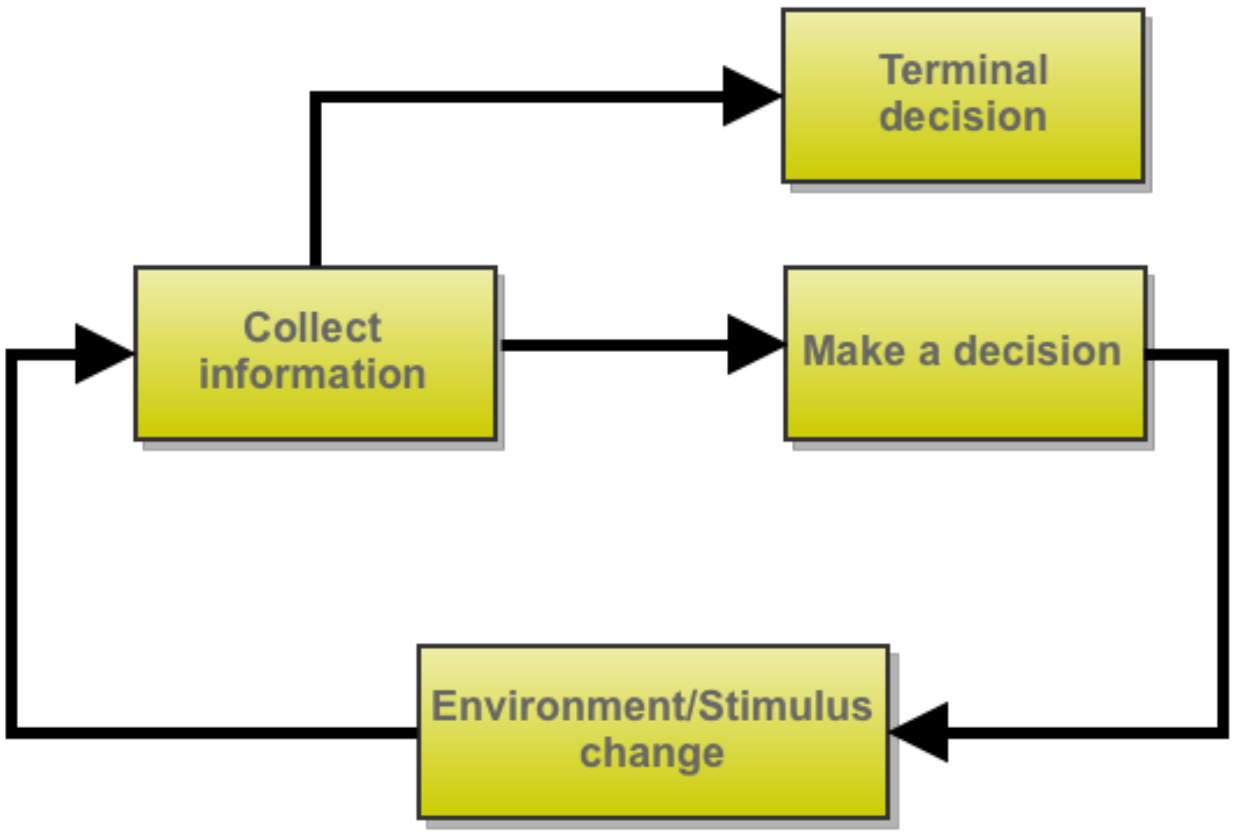
Low Dimensional Approximate Control
- Results

Comparison with Infomax Policy

Conclusion

Introduction

- Active sensing falls under the more general area of closed loop decision making.
- The underlying problem structure being:



Introduction

- Other examples of such decision making problems include:
  - Sensor management [Hero and Cochran, 2011]
  - Generalized binary search [Nowak, 2011]
  - Teaching word meanings [Whitehill and Movellan, 2012]
  - Underwater object classification [Hollinger et al., 2011]
  - Menu design for P300 prosthetic [Jarzebowski et al., 2012]
- A natural framework to study these problems is Markov Decision Processes (MDPs) or Partially Observable Markov Decision Processes (POMDPs).
- Exact solutions are computationally inefficient especially for POMDPs.
- General as well as application specific approximations are an active research field [Powell, 2007; Lagoudakis and Parr, 2003; Kaplow, 2010].

Outline

Introduction

Active Sensing: Background

Visual Search Task

POMDP Formulation

- Bayesian Inference
- Optimal Action Selection
- Results
- Scalability Issues

Low Dimensional Approximate Control

- Results

Comparison with Infomax Policy

Conclusion

Active Sensing: Background

- Problem of choosing fixation location has been well-studied.
- Feedforward approaches include random fixations, using saliency maps [Itti et al., 1998], fixating class separating locations [Lacroix et al., 2008], etc.
- Usually very simple, and describe some free viewing behavior.
- Some shortcomings:
  - Lack of provision to query peripheral locations.
  - Lack of inherent mechanism to implement inhibition of return.
  - Saliency has been shown to play little role in goal-oriented visual tasks [Yarbus, 1967].

Active Sensing: Background

- Feedback approaches include maximizing one step detection probability [Najemnik and Geisler, 2005], minimizing entropy [Butko and Movellan, 2010], etc.
- Such surrogate goals can yield computationally tractable policies, with some performance guarantees [Williams et al., 2007].
- Some shortcomings:
  - Lack of provision for task specific demands or behavioral costs.
  - Require an ad-hoc stopping criteria for terminal decision.
  - More descriptive than predictive.
- Ideal goal: Computationally tractable policy that also overcomes these shortcomings.
- Contribution: Solve for exact optimal policy that explains human data; use the insights gained to design approximations, and to augment existing algorithms.

Outline

Introduction

Active Sensing: Background

Visual Search Task

POMDP Formulation
- Bayesian Inference
- Optimal Action Selection
- Results
- Scalability Issues

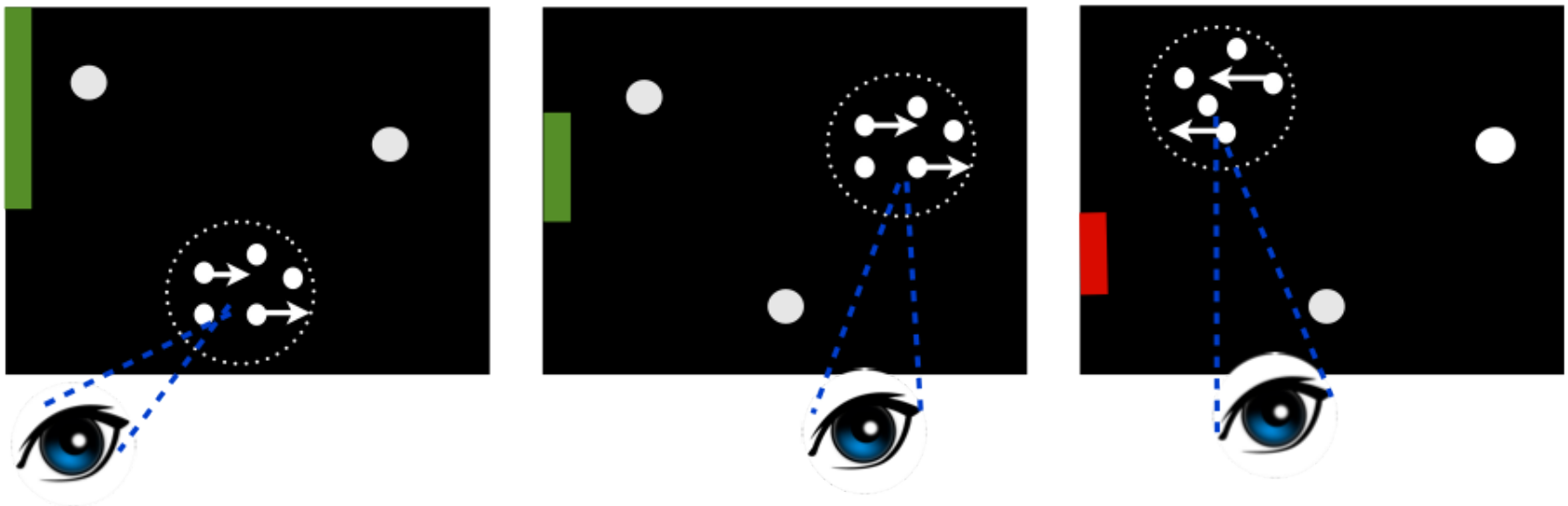
Low Dimensional Approximate Control
- Results

Comparison with Infomax Policy

Conclusion

Visual Search Task

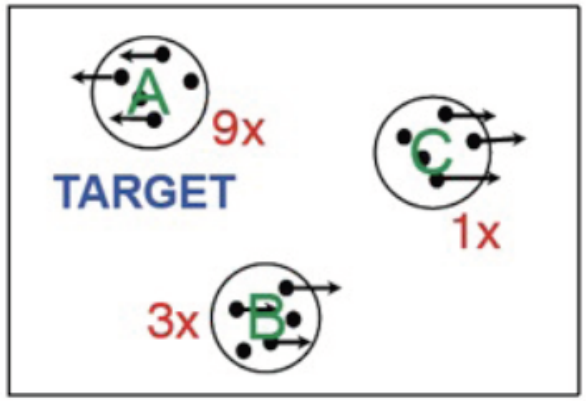
[Huang and Yu, SfN, 2010]



- Task: Find the target ($\Leftarrow$) amongst the distractors ($\Rightarrow$).
- Gaze contingent display allows exact measurement of where subject obtains sensory input.
- Sequence of stimulus controlled by subject.

Visual Search Task

- Some locations more likely to be the target (1:3:9)



- Reward policy:

$$\text{Score (\$)} = 100 - 12.5 \times T - 25 \times S + 50 \times C$$

base points

search cost

switch cost

accuracy

Outline

Introduction

Active Sensing: Background

Visual Search Task

POMDP Formulation
- Bayesian Inference
- Optimal Action Selection
- Results
- Scalability Issues

Low Dimensional Approximate Control
- Results

Comparison with Infomax Policy

Conclusion

POMDP Formulation

- Loss formulation of a POMDP is a six-tuple (S, A, O, T, Ω, L).

POMDP Formulation

- Loss formulation of a POMDP is a six-tuple (S, A, O, T, Ω, L).
- S (set of states): Set of target locations $\{1, 2, 3\}$.

POMDP Formulation

- Loss formulation of a POMDP is a six-tuple (S, A, O, T, Ω, L).
- S (set of states): Set of target locations $\{1, 2, 3\}$.
- A (set of actions): Next location to fixate $\{1, 2, 3\}$, and terminal (stopping) action $\{0\}$.

POMDP Formulation

- Loss formulation of a POMDP is a six-tuple (S, A, O, T, Ω, L).
- S (set of states): Set of target locations $\{1, 2, 3\}$.
- A (set of actions): Next location to fixate $\{1, 2, 3\}$, and terminal (stopping) action $\{0\}$.
- O (set of observations): Direction of dots $\{0(\textit{right}), 1(\textit{left})\}$.

POMDP Formulation

- Loss formulation of a POMDP is a six-tuple (S, A, O, T, Ω, L).
- S (set of states): Set of target locations $\{1, 2, 3\}$.
- A (set of actions): Next location to fixate $\{1, 2, 3\}$, and terminal (stopping) action $\{0\}$.
- O (set of observations): Direction of dots $\{0(\textit{right}), 1(\textit{left})\}$.
- T (set of transition probabilities): A $3x3$ identity matrix.

POMDP Formulation

- Loss formulation of a POMDP is a six-tuple (S, A, O, T, Ω, L).
- S (set of states): Set of target locations $\{1, 2, 3\}$.
- A (set of actions): Next location to fixate $\{1, 2, 3\}$, and terminal (stopping) action $\{0\}$.
- O (set of observations): Direction of dots $\{0(\textit{right}), 1(\textit{left})\}$.
- T (set of transition probabilities): A $3x3$ identity matrix.
- Ω (set of observation probabilities):

$$\Omega(o|s, a) = \mathbb{1}_{\{s=a\}}\ \textit{Bern}(o, \beta) + \mathbb{1}_{\{s\neq a\}}\ \textit{Bern}(o, 1-\beta)$$

POMDP Formulation

- Loss formulation of a POMDP is a six-tuple (S, A, O, T, Ω, L).
- S (set of states): Set of target locations $\{1, 2, 3\}$.
- A (set of actions): Next location to fixate $\{1, 2, 3\}$, and terminal (stopping) action $\{0\}$.
- O (set of observations): Direction of dots $\{0(\textit{right}), 1(\textit{left})\}$.
- T (set of transition probabilities): A $3x3$ identity matrix.
- Ω (set of observation probabilities):

$$\Omega(o|s, a) = \mathbb{1}_{\{s=a\}}\ \textit{Bern}(o, \beta) + \mathbb{1}_{\{s\neq a\}}\ \textit{Bern}(o, 1-\beta)$$

- L (loss function):

$$L(s, a_{t-1}, a_t) = \begin{cases} \mathbb{1}_{\{s\neq a_{t-1}\}} & \text{if } a_t = 0 \\ c + c_s \mathbb{1}_{\{a_t\neq a_{t-1}\}} & \text{if } a_t \in \{1, 2, 3\} \end{cases}$$

where c is cost of unit time and c_s is cost of a switch.

Bayesian Inference

- The agent does not know the exact state (target location).
- Instead it maintains a probability distribution on states, known as belief states:

$$\mathbf{b}_t = \big((p(s=1|\mathbf{o}_t;\mathbf{a}_t),(p(s=2|\mathbf{o}_t;\mathbf{a}_t),(p(s=3|\mathbf{o}_t;\mathbf{a}_t)\big)$$

  where $\mathbf{o}_t$ is observation history, and $\mathbf{a}_t$ is fixation location history till time t.
- Belief update using Bayes rule:

$$\mathbf{b}_t(s) \propto p(o_t|s;a_t)\ p(s|\mathbf{o}_{t-1};\mathbf{a}_{t-1}) = \Omega(o_t|s,a_t)\mathbf{b}_{t-1}(s)$$

Optimal Action Selection

- A policy (π) is a function mapping belief states to actions.
- The value of a policy is defined as the expected loss incurred following that policy:

$$V^{\pi}(\mathbf{b}_t, a_t) = \sum_{t'=t+1}^{\infty} \mathbb{E}[L_{t'} | \mathbf{b}_{t'}, \pi]$$

- The optimal policy is thus:

$$\pi^*(\mathbf{b}_t, a_t) = \underset{\pi}{\operatorname{argmin}}\, V^{\pi}(\mathbf{b}_t, a_t)$$

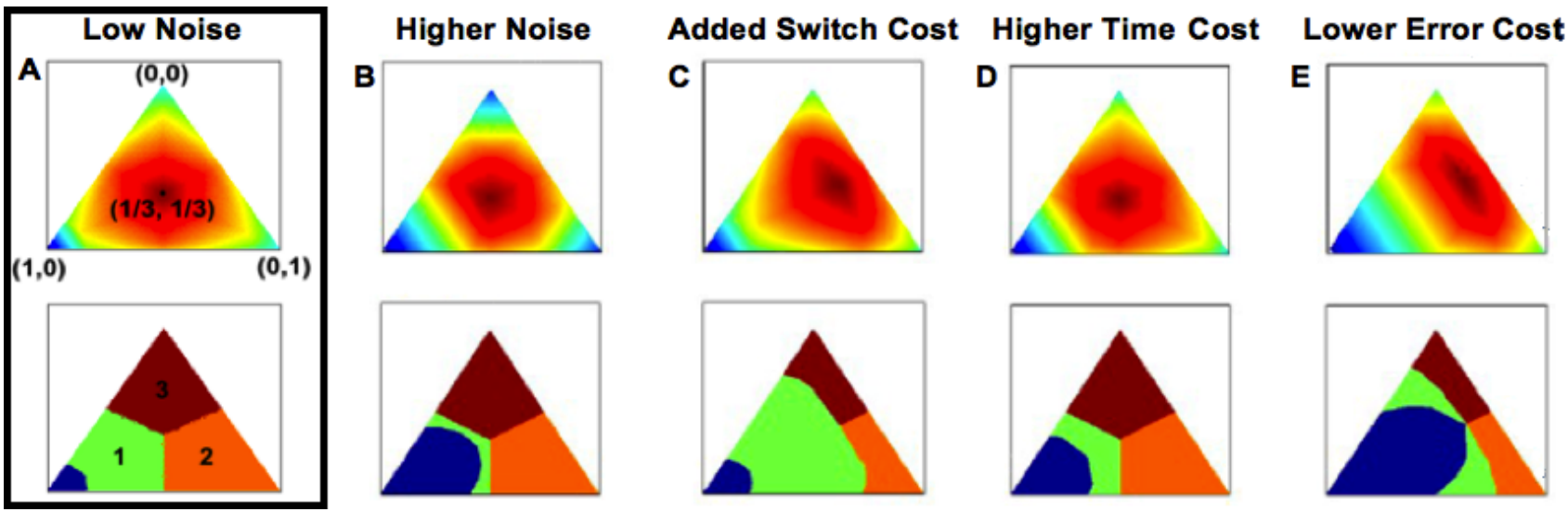
- Bellman optimality equation [Bellman, 1952]:

$$V^*(\mathbf{b}_t, a_t) = \min \begin{cases} (1 - \mathbf{b}_t(a_t)) & \text{if } a_{t+1} = 0 \\ c + c_s \mathbb{1}_{\{a_{t+1} \neq a_t\}} + \mathbb{E}[V^*(\mathbf{b}_{t+1}, a_{t+1})] & \text{otherwise} \end{cases}$$

Results: Optimal Policy

- Results shown over a gridded belief state (size = 201).
- Grid-based approximation improves with grid density [Lovejoy, 1991], but computationally inefficient.

Environment $(c, c_s, \beta) = (0.1, 0, 0.9)$

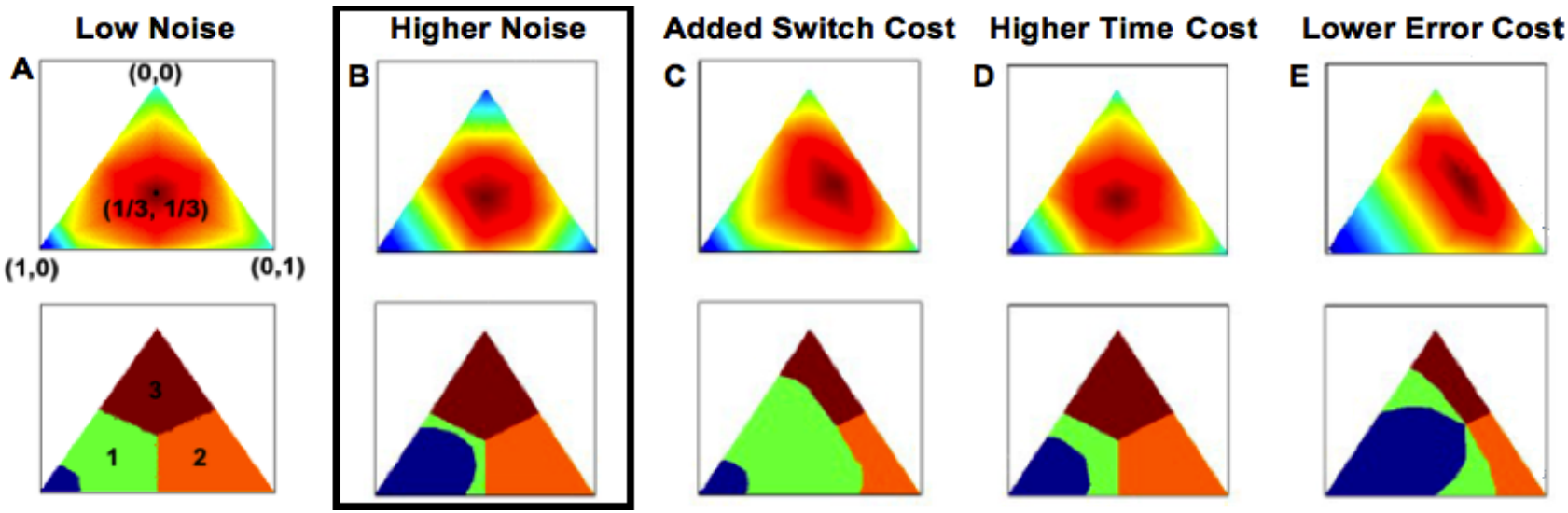


Stop at high certainty

Results: Optimal Policy

- Results shown over a gridded belief state (size = 201).
- Grid-based approximation improves with grid density [Lovejoy, 1991], but computationally inefficient.

Environment $(c, c_s, \beta) = (0.1, 0, 0.7)$

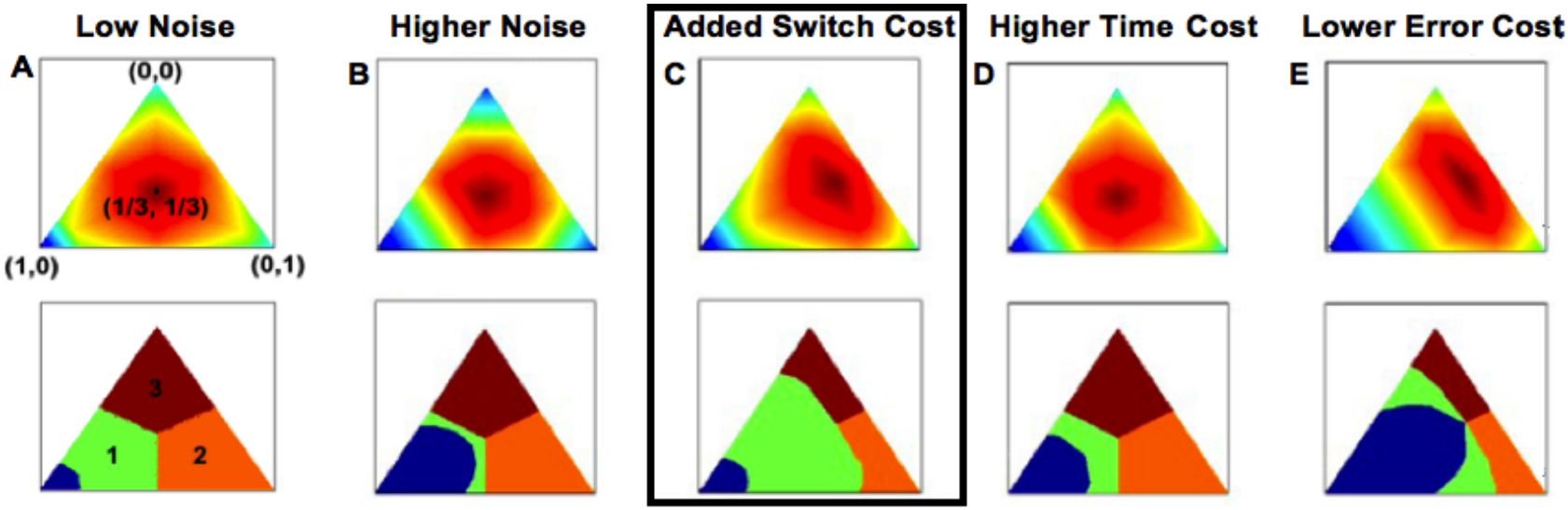


Stop early

Results: Optimal Policy

- Results shown over a gridded belief state (size = 201).
- Grid-based approximation improves with grid density [Lovejoy, 1991], but computationally inefficient.

Environment $(c, c_s, \beta) = (0.1, 0.1, 0.9)$

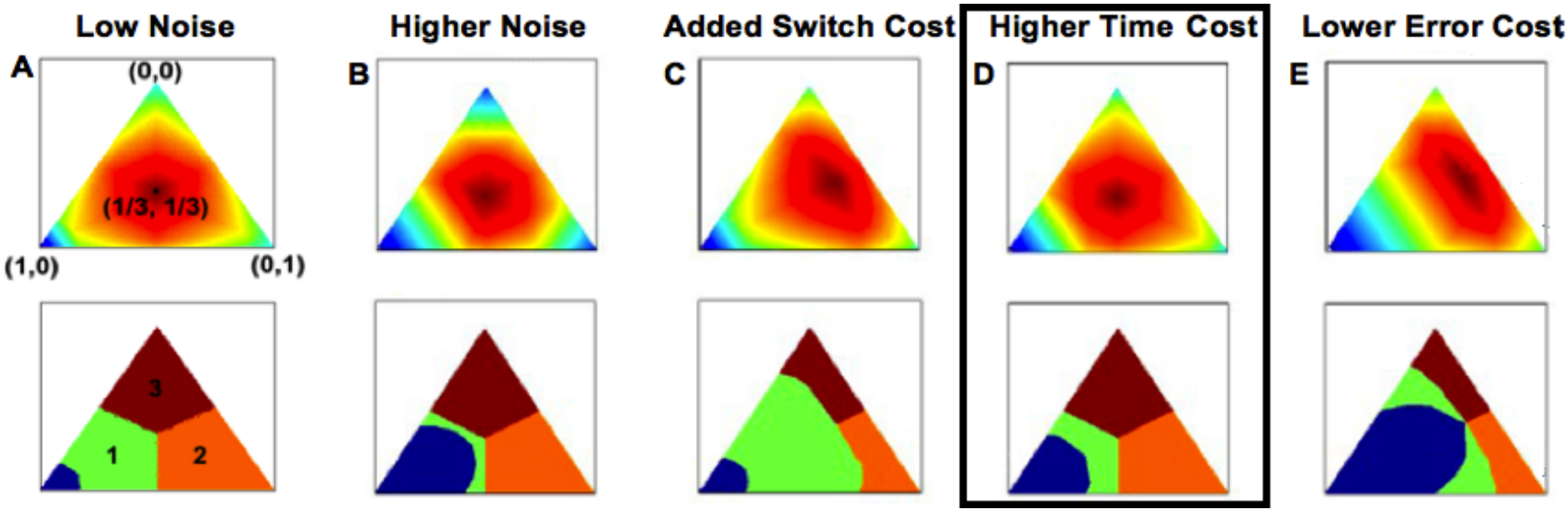


Switch less

Results: Optimal Policy

- Results shown over a gridded belief state (size = 201).
- Grid-based approximation improves with grid density [Lovejoy, 1991], but computationally inefficient.

Environment $(c, c_s, \beta) = (0.2, 0, 0.9)$

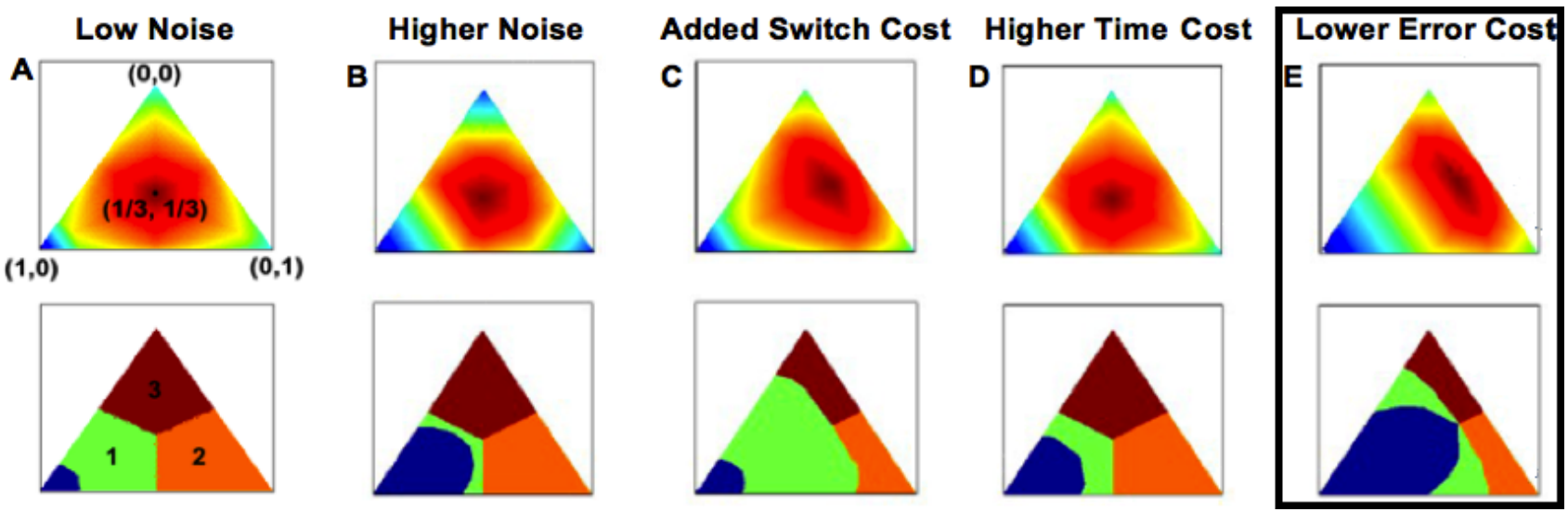


Stop early

Results: Optimal Policy

- Results shown over a gridded belief state (size = 201).
- Grid-based approximation improves with grid density [Lovejoy, 1991], but computationally inefficient.

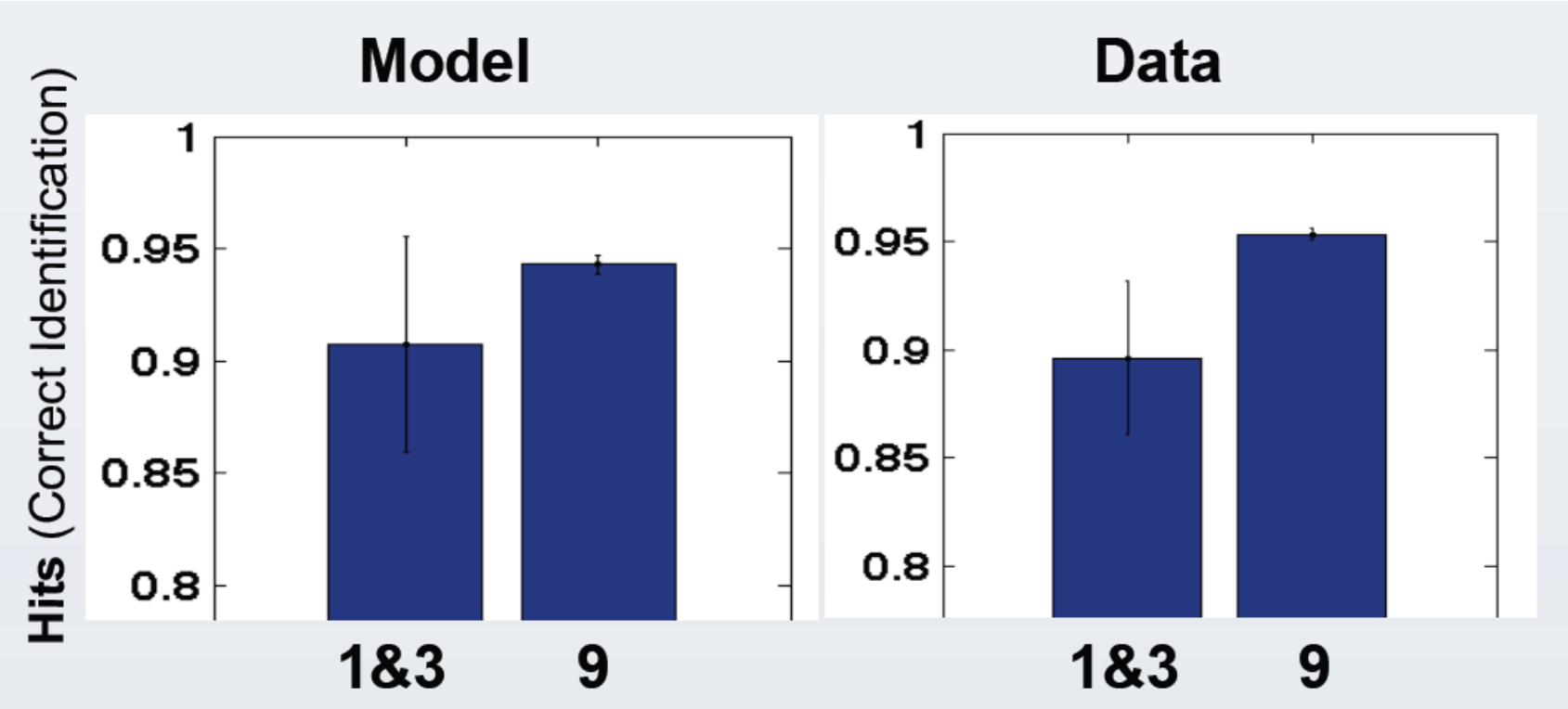
Environment $(c, c_s, \beta) = (0.2, 0.2, 0.9)$



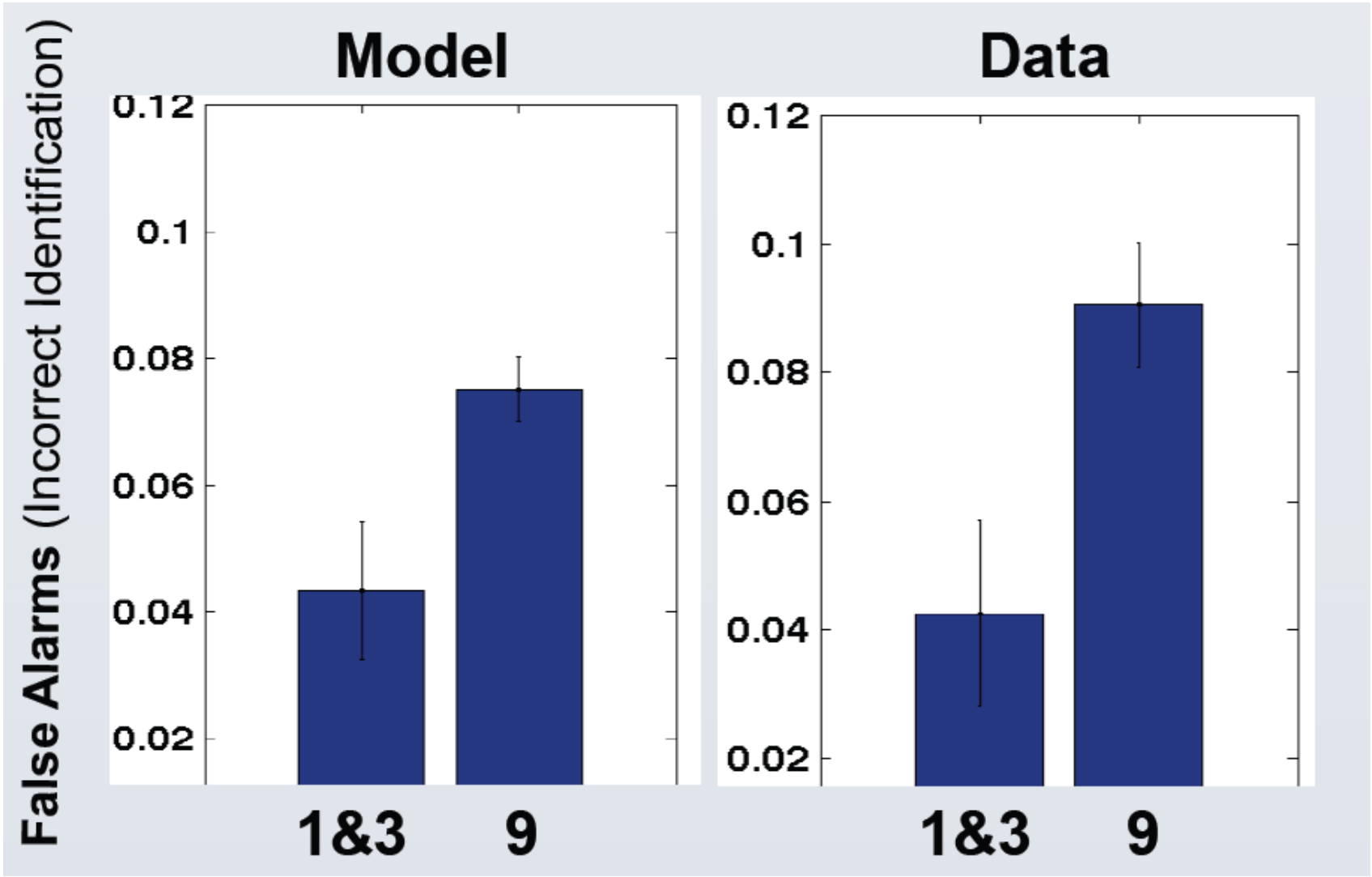
Stop early, switch less

Results: Confirmation Bias I

$\uparrow$ prior expectation $\Rightarrow \uparrow P(\text{target selection})$

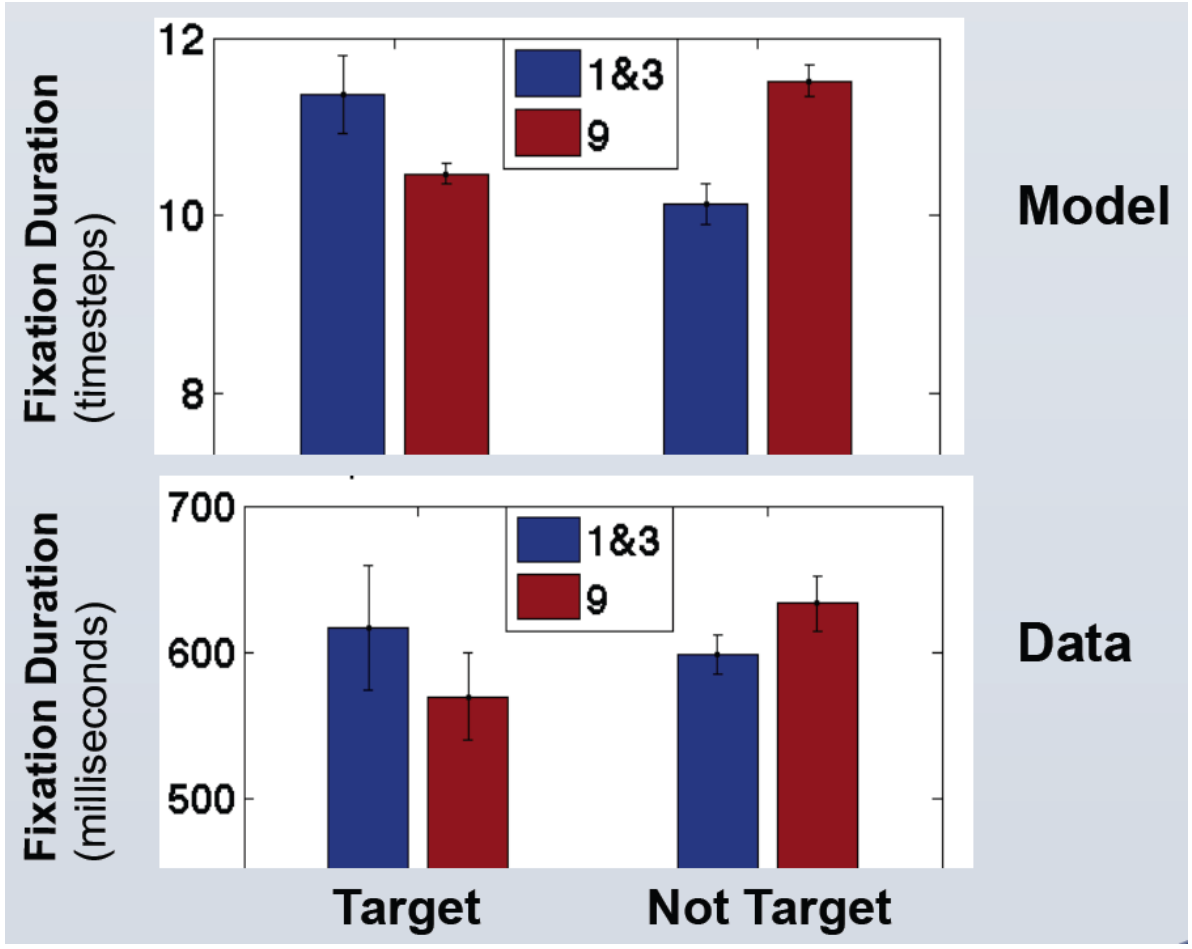


Results: Confirmation Bias II



Results: Confirmation Bias III

$\uparrow$ prior expectation $\Rightarrow \downarrow$ time to confirm, $\uparrow$ time to disconfirm



Scalability Issues

- Belief state MDP formulations suffer from the *curse of dimensionality*.
- The state-space (belief state) is continuous, hence infinite dimensional.
- Algorithmic complexity is $O(kn^{k-1})$, for k sensing locations and a grid-size of n.
- Next we present simpler approximations (complexity linear in k), that also retain context sensitivity.

Outline

Introduction

Active Sensing: Background

Visual Search Task

POMDP Formulation

- Bayesian Inference
- Optimal Action Selection
- Results
- Scalability Issues

Low Dimensional Approximate Control

- Results

Comparison with Infomax Policy

Conclusion

Low Dimensional Approximate Control

1. Fix M Radial Basis Functions (RBF): $\phi(\mathbf{b}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{b}-\mu_j||^2}{2\sigma^2}}$

Low Dimensional Approximate Control

1. Fix M Radial Basis Functions (RBF): $\phi(\mathbf{b}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{b}-\mu_i||^2}{2\sigma^2}}$
2. Generate m points randomly from the belief space ($\mathbf{b}$).

Low Dimensional Approximate Control

1. Fix M Radial Basis Functions (RBF): $\phi(\mathbf{b}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{b}-\mu_i||^2}{2\sigma^2}}$
2. Generate m points randomly from the belief space ($\mathbf{b}$).
3. Initialize the value function($\{V(\mathbf{b}_i)\}_{i=1}^{m}$) with the stopping costs.

Low Dimensional Approximate Control

1. Fix M Radial Basis Functions (RBF): $\phi(\mathbf{b}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{b}-\mu_i||^2}{2\sigma^2}}$
2. Generate m points randomly from the belief space ($\mathbf{b}$).
3. Initialize the value function($\{V(\mathbf{b}_i)\}_{i=1}^{m}$) with the stopping costs.
4. Find $\mathbf{w}$, the minimum norm solution of: $V(\mathbf{b}) = \Phi(\mathbf{b})\mathbf{w}$.

Low Dimensional Approximate Control

1. Fix M Radial Basis Functions (RBF): $\phi(\mathbf{b}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{b}-\mu_i||^2}{2\sigma^2}}$
2. Generate m points randomly from the belief space ($\mathbf{b}$).
3. Initialize the value function($\{V(\mathbf{b}_i)\}_{i=1}^{m}$) with the stopping costs.
4. Find $\mathbf{w}$, the minimum norm solution of: $V(\mathbf{b}) = \Phi(\mathbf{b})\mathbf{w}$.
5. Generate a new set of m random belief state points ($\mathbf{b}'$).

Low Dimensional Approximate Control

1. Fix M Radial Basis Functions (RBF): $\phi(\mathbf{b}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{b}-\mu_i||^2}{2\sigma^2}}$
2. Generate m points randomly from the belief space ($\mathbf{b}$).
3. Initialize the value function($\{V(\mathbf{b}_i)\}_{i=1}^{m}$) with the stopping costs.
4. Find $\mathbf{w}$, the minimum norm solution of: $V(\mathbf{b}) = \Phi(\mathbf{b})\mathbf{w}$.
5. Generate a new set of m random belief state points ($\mathbf{b}'$).
6. Evaluate required V values for value iteration using current $\mathbf{w}$.

Low Dimensional Approximate Control

1. Fix M Radial Basis Functions (RBF): $\phi(\mathbf{b}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{b}-\mu_i||^2}{2\sigma^2}}$
2. Generate m points randomly from the belief space ($\mathbf{b}$).
3. Initialize the value function($\{V(\mathbf{b}_i)\}_{i=1}^{m}$) with the stopping costs.
4. Find $\mathbf{w}$, the minimum norm solution of: $V(\mathbf{b}) = \Phi(\mathbf{b})\mathbf{w}$.
5. Generate a new set of m random belief state points ($\mathbf{b}'$).
6. Evaluate required V values for value iteration using current $\mathbf{w}$.
7. Update $V(\mathbf{b}')$ using value iteration.

Low Dimensional Approximate Control

1. Fix M Radial Basis Functions (RBF): $\phi(\mathbf{b}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{b}-\mu_i||^2}{2\sigma^2}}$
2. Generate m points randomly from the belief space ($\mathbf{b}$).
3. Initialize the value function($\{V(\mathbf{b}_i)\}_{i=1}^{m}$) with the stopping costs.
4. Find $\mathbf{w}$, the minimum norm solution of: $V(\mathbf{b}) = \Phi(\mathbf{b})\mathbf{w}$.
5. Generate a new set of m random belief state points ($\mathbf{b}'$).
6. Evaluate required V values for value iteration using current $\mathbf{w}$.
7. Update $V(\mathbf{b}')$ using value iteration.
8. Find a new $\mathbf{w}$ from $V(\mathbf{b}') = \Phi(\mathbf{b}')\mathbf{w}$.

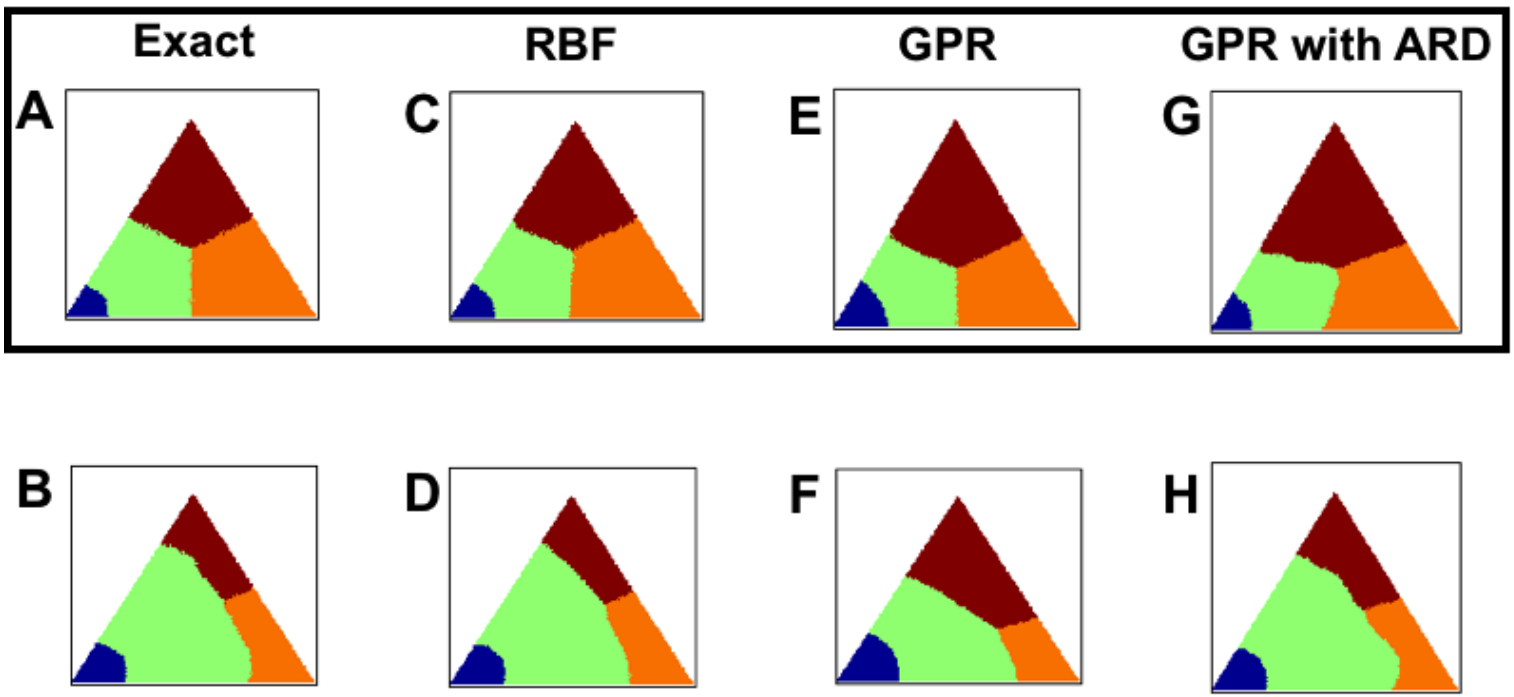
Low Dimensional Approximate Control

1. Fix M Radial Basis Functions (RBF): $\phi(\mathbf{b}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{b}-\mu_i||^2}{2\sigma^2}}$
2. Generate m points randomly from the belief space ($\mathbf{b}$).
3. Initialize the value function($\{V(\mathbf{b}_i)\}_{i=1}^{m}$) with the stopping costs.
4. Find $\mathbf{w}$, the minimum norm solution of: $V(\mathbf{b}) = \Phi(\mathbf{b})\mathbf{w}$.
5. Generate a new set of m random belief state points ($\mathbf{b}'$).
6. Evaluate required V values for value iteration using current $\mathbf{w}$.
7. Update $V(\mathbf{b}')$ using value iteration.
8. Find a new $\mathbf{w}$ from $V(\mathbf{b}') = \Phi(\mathbf{b}')\mathbf{w}$.
9. Repeat steps 5 through 8, until w converges.

Results: Comparison with Approximate Policies

- Results shown for RBF, Gaussian Processes Regression (GPR) [Williams and Rasmussen, 1996] and GPR with Automatic Relevance Determination (ARD).
- Grid size $= 201$. RBF: $M = 49$, $m = 1000$.

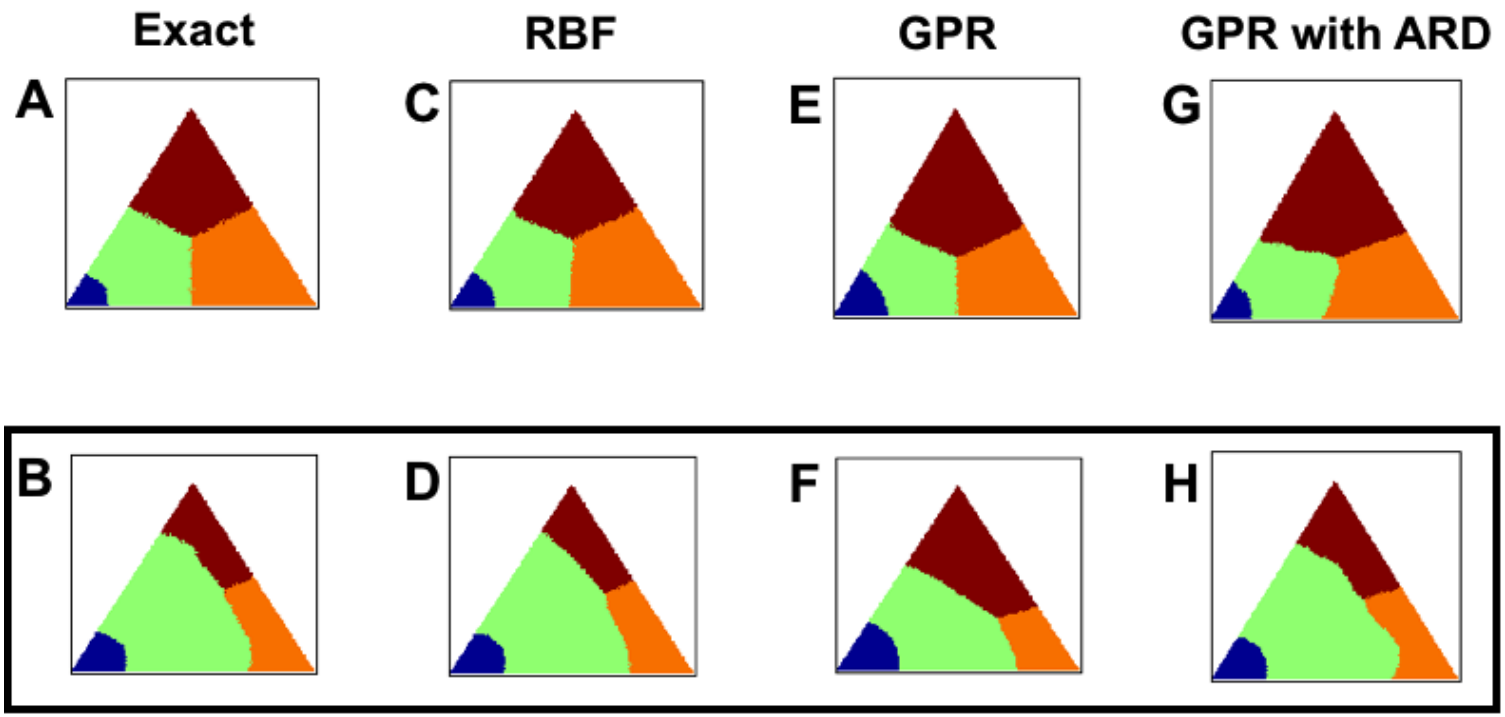
Environment $(c, c_s, \beta) = (0.1, 0, 0.9)$



Results: Comparison with Approximate Policies

- Results shown for RBF, Gaussian Processes Regression (GPR) [Williams and Rasmussen, 1996] and GPR with Automatic Relevance Determination (ARD).
- Grid size $= 201$. RBF: $M = 49$, $m = 1000$.

Environment $(c, c_s, \beta) = (0.1, 0.1, 0.9)$



Outline

Introduction

Active Sensing: Background

Visual Search Task

POMDP Formulation

- Bayesian Inference
- Optimal Action Selection
- Results
- Scalability Issues

Low Dimensional Approximate Control

- Results

Comparison with Infomax Policy

Conclusion

Comparison with Infomax Policy

- Infomax [Butko and Movellan, 2010] also tackles a visual search problem.
- Uses finite horizon entropy as the cost function.
- Insights gained from the geometry of optimal policy can be used to parametrically augment Infomax policy.

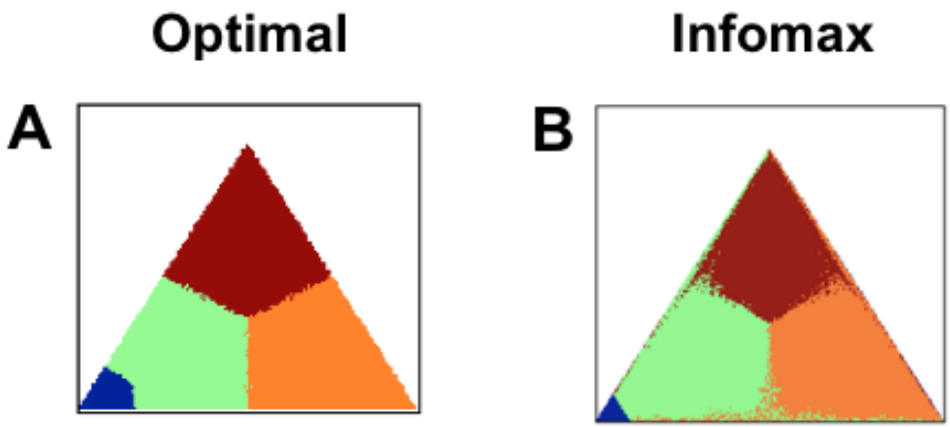


Figure: *Policies shown over* 201 *bins.* $c = 0.1, c_s = 0, \beta = 0.9$. *(A) Behavioral policy. (B) Infomax policy (stop when posterior belief exceeds 0.9)*

Outline

Introduction

Active Sensing: Background

Visual Search Task

POMDP Formulation

- Bayesian Inference
- Optimal Action Selection
- Results
- Scalability Issues

Low Dimensional Approximate Control

- Results

Comparison with Infomax Policy

Conclusion

Conclusion

- Presented an active sensing framework that takes into account task demands and behavioral costs.
- Application to a simple visual search task makes intuitive predictions.
- Comparison with human data shows close fit, explains confirmation bias.
- Presented approximate algorithms that are computationally tractable yet context sensitive.
- The work aims to add to the growing literature on problems in decision processes, to sprout new approximations and to augment existing algorithms.
- We believe that a framework sensitive towards behavioral costs can not only lead to better artificial agents, but also provide us with neurological underpinnings of active sensing.

References I

R Bellman. On the theory of dynamic programming. *PNAS*, 38(8):716–719, 1952.

N J Butko and J R Movellan. Infomax control of eyemovements. *IEEE Transactions on Autonomous Mental Development*, 2(2):91–107, 2010.

A.O. Hero and D. Cochran. Sensor management: Past, present, and future. *Sensors Journal, IEEE*, 11(12):3064 –3075, December 2011. ISSN 1530-437X. doi: 10.1109/JSEN.2011.2167964.

G.A. Hollinger, U. Mitra, and G.S. Sukhatme. Active classification: Theory and application to underwater inspection. *arXiv preprint arXiv:1106.5829*, 2011.

L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11):1254–1259, 1998.

J. Jarzebowski, R. Ma, N. Aghasadeghi, T. Bretl, and T.P. Coleman. A stochastic control approach to optimally designing variable-sized menus in p300 communication prostheses. 2012.

R. Kaplow. *Point-based POMDP solvers: Survey and comparative analysis*. PhD thesis, McGill University, 2010.

J. Lacroix, E. Postma, J. Van Den Herik, and J. Murre. Toward a visual cognitive system using active top-down saccadic control. *International Journal of Humanoid Robotics*, 5(02):225–246, 2008.

M.G. Lagoudakis and R. Parr. Least-squares policy iteration. *The Journal of Machine Learning Research*, 4:1107–1149, 2003.

References II

W.S. Lovejoy. Computationally feasible bounds for partially observed markov decision processes. *Operations research*, 39(1):162–175, 1991.

J Najemnik and W S Geisler. Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–91, 2005.

R.D. Nowak. The geometry of generalized binary search. *Information Theory, IEEE Transactions on*, 57(12):7893–7906, 2011.

W.B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. Wiley-Interscience, 2007.

J. Whitehill and J. Movellan. Teaching word meanings by visual examples. *Journal of Machine Learning Research*, 2012.

C K I Williams and C E Rasmussen. Gaussian processes for regression. In M.C. Mozer D. S. Touretzky and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, pages 514–20. MIT Press, Cambridge, MA, 1996.

J.L. Williams, J.W. Fisher III, and A.S. Willsky. Performance guarantees for information theoretic active inference. *AI & Statistics (AISTATS)*, 2007.

A F Yarbus. *Eye Movements and Vision*. Plenum Press, New York, 1967.

Thanks!!

Additional Slides

- Complexity of RBF approximation is $O(k(mM + M^3))$
- Complexity of GPR approximation is $O(kN^3)$, where N is the number of points used for regression.
- For GPR simulations: 200 points used for extrapolation at each step, length scale $= 1$, signal strength $= 1$ and noise strength $= 0.1$
- Approximation motivated by Warren Powell's book [Powell, 2007] and LSPI [Lagoudakis and Parr, 2003].