Optimal Continuous State Planning with Semantic Observations

Luke Burks
Graduate Research Assistant

Nisar Ahmed
Assistant Professor

Conference on Decision and Control, Melbourne, VIC, Australia

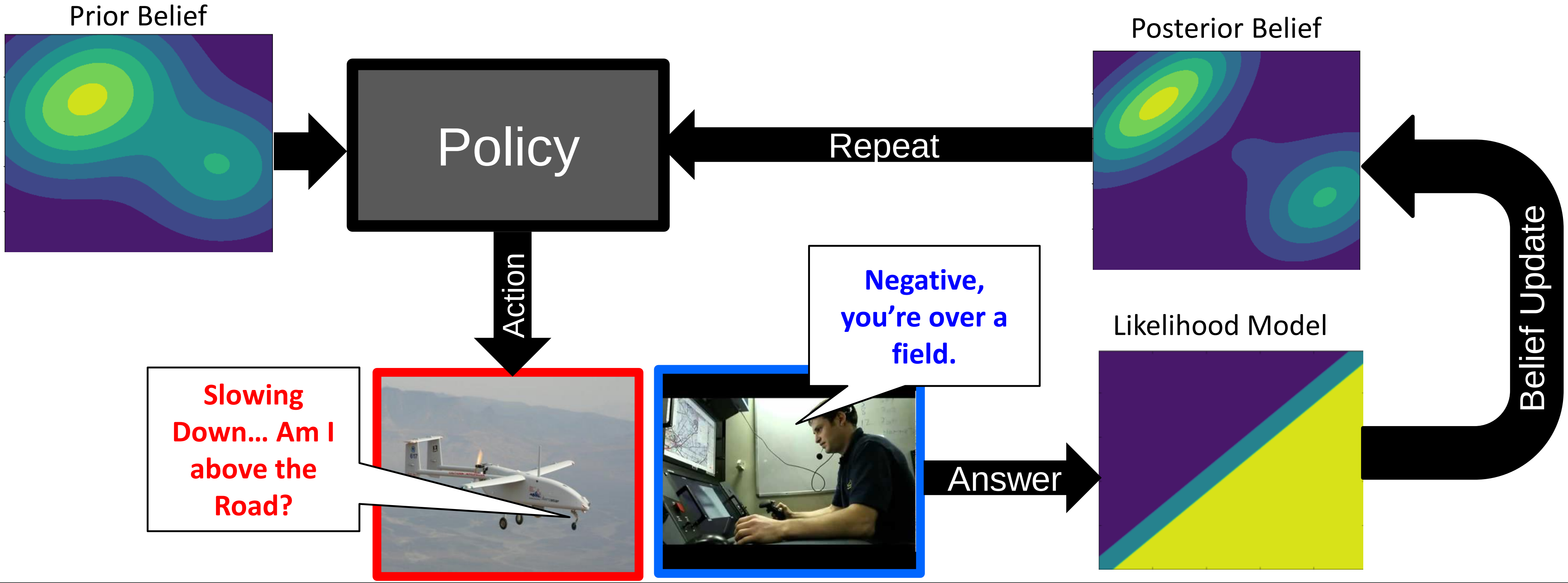
December 12, 2017

Cooperative Human-Robot Intelligence Laboratory
Ann and H.J. Smead Aerospace Engineering Sciences
University of Colorado at Boulder

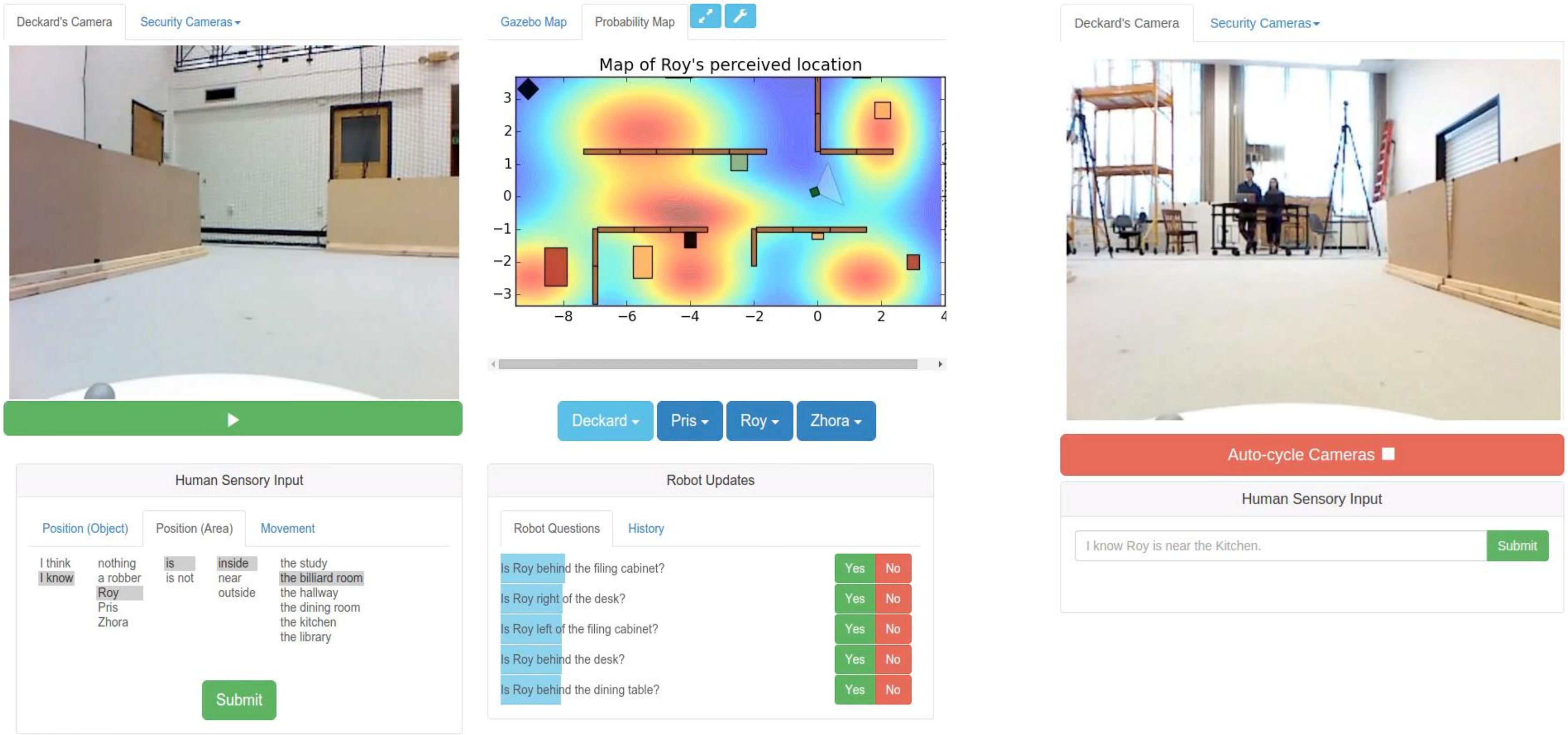
COHRINT

Research Vision

- Treat humans as taskable information providers for autonomous robots
- Formally <u>integrate</u> semantic human observations into <u>tightly coupled optimal</u> sensing and planning under uncertainty

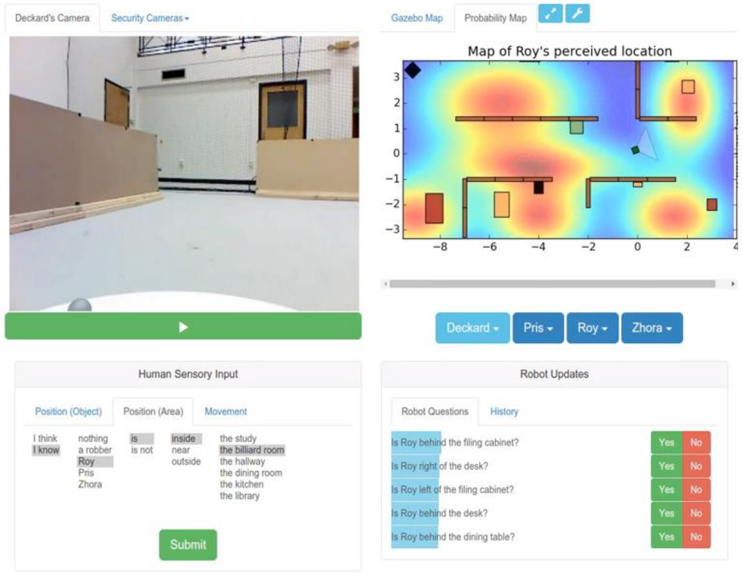


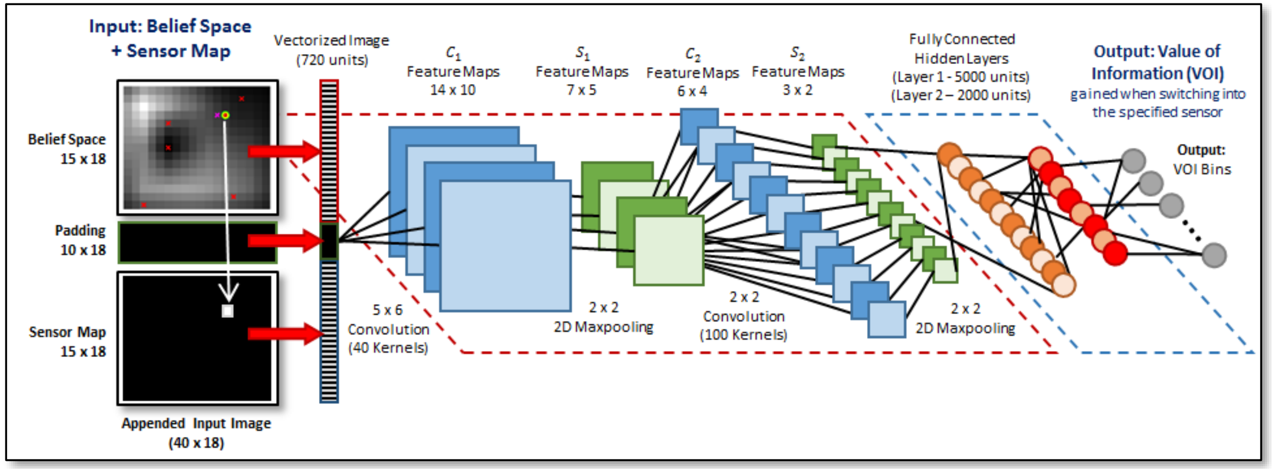
Cops and Robots Platform



Related Work and Issues: Semantic Sensing and Planning

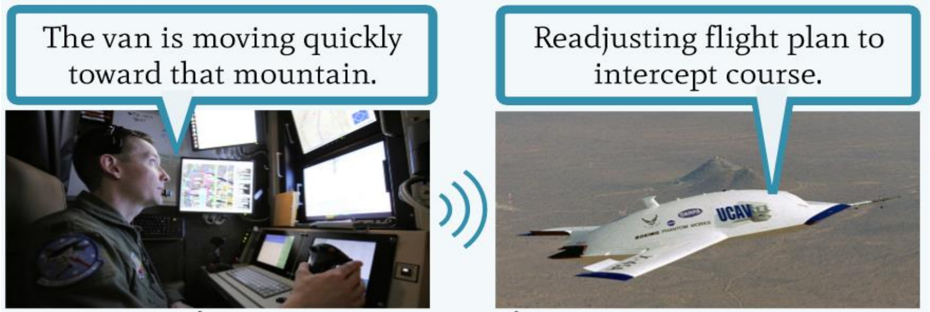
- Decoupled Planning and Control
  [Sweet, Ahmed, ACC 2016]
  - Doesn't account for information gathering actions
- Discretized State Space
  [Kurniawati, Hsu, Lee, RSS 2008]
  - Difficult to scale to larger state spaces
- Deep Learning
  [Lore, et al., ICCPS 2016]
  - Requires large amounts of training data
- Online POMDP
  [Silver, Veness, NIPS 2010]
  - Brittle in scenarios without intermediate rewards
- CPOMPS using Gaussian Mixtures
  [Porta, et al., IJCAI 2011]
  - Avoids issues above
  - Difficult to specify observation models with GMs, GM explosion

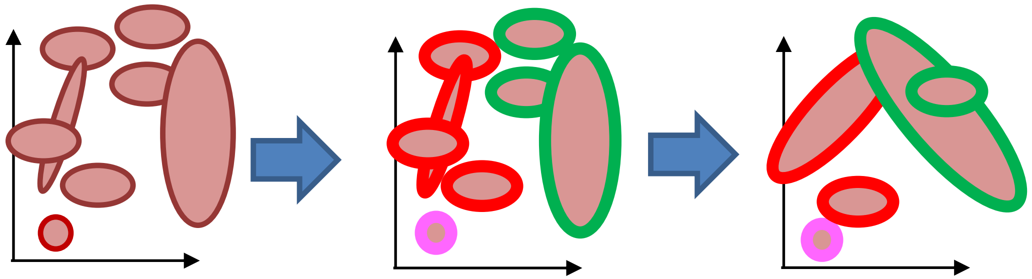




Contributions of this Work

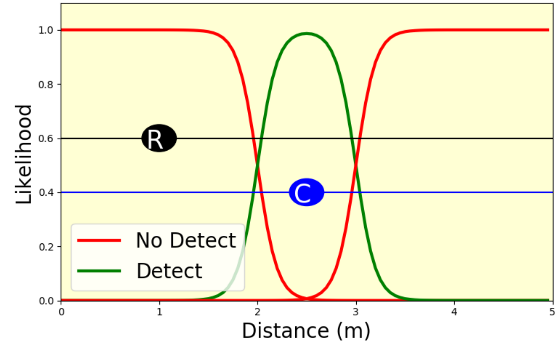
- Semantic observation modeling in continuous state POMDPs (CPOMDPs)
  - Gaussian Mixture (GM) policy functions for softmax observation models using Variational Bayes





- Novel method for taming exponential GM explosion
  - Fast pre-clustering + condensation within clusters
  - Parallelized the merging process

- Proof of Concept in Simulation
  - Better scaling for large observation spaces
  - More efficient policy updates → Faster convergence



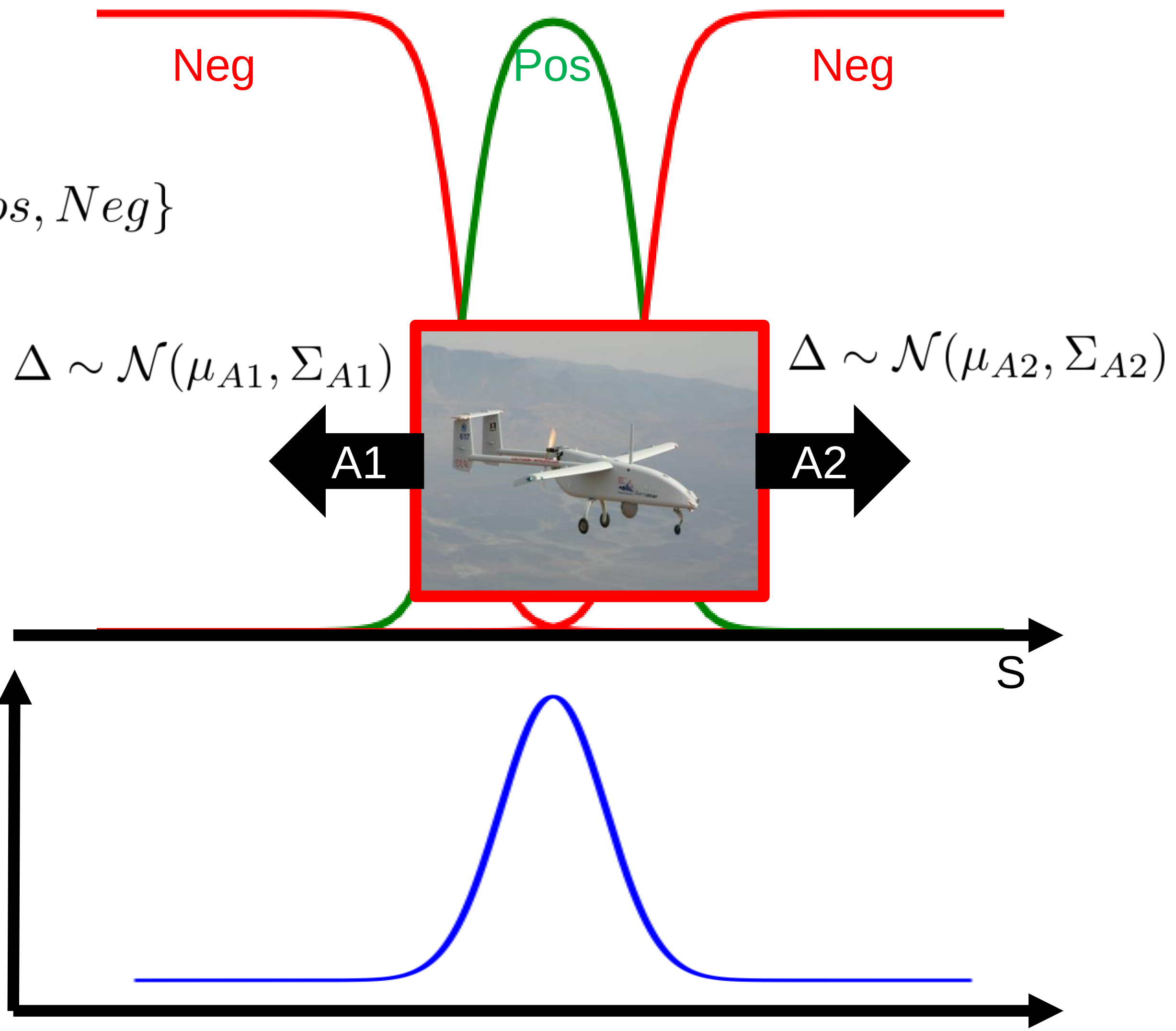
Continuous POMDPs (CPOMDPs)

$State = \mathbf{S} = (-\infty, \infty)$

Unknown target state, known searcher state

$Observations = \Omega = \{Pos, Neg\}$

$Actions = \{A1, A2\}$



- Rewards given for being near the target

How to choose actions?

- POMDP solvers find policies to map beliefs to actions:

$$\pi(b) \rightarrow a$$

- Policies maximize discounted expected reward over time:

$$E[\sum_{t=0}^{\infty} \gamma^t r_t]$$

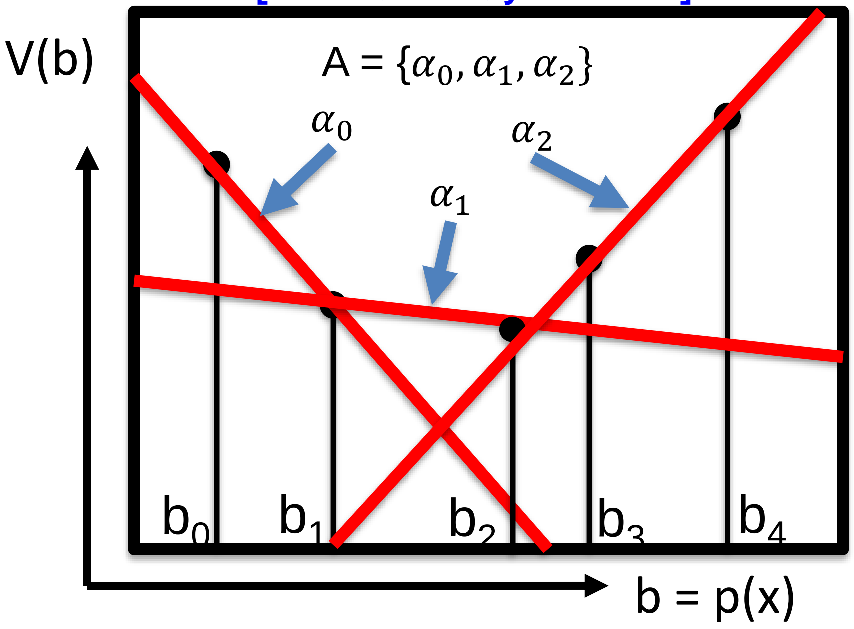
CPOMDPs: Approximate Value Iteration in Belief Space

Alpha Element Backup for Point-Based Value Iteration

$$\alpha^i_{a,o}(s) = \int_{s'} \alpha^i_{n-1}(s')p(o|s')p(s'|s,a)ds'$$

$$\alpha^i_n(s) = r_a(s) + \gamma \sum_o \arg\max_{\alpha^i_{a,o}}(<\alpha^i_{a,o}, b>)$$

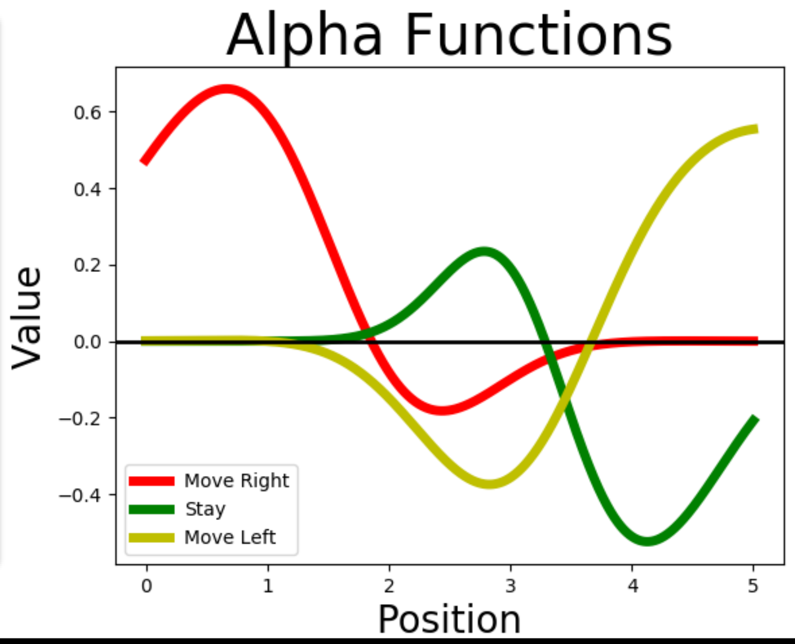
PBVI-type solution on discretized space x
with α-vectors for policy π
[Pineau, et al., JAIR 2006]

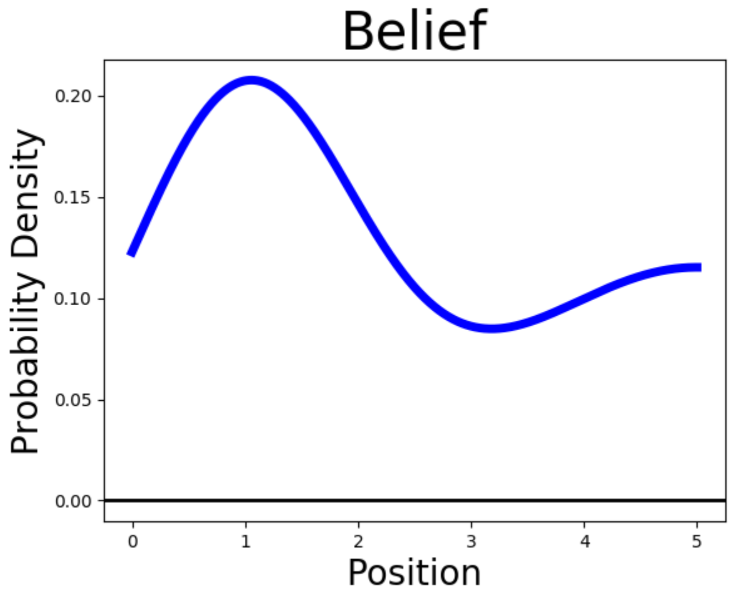


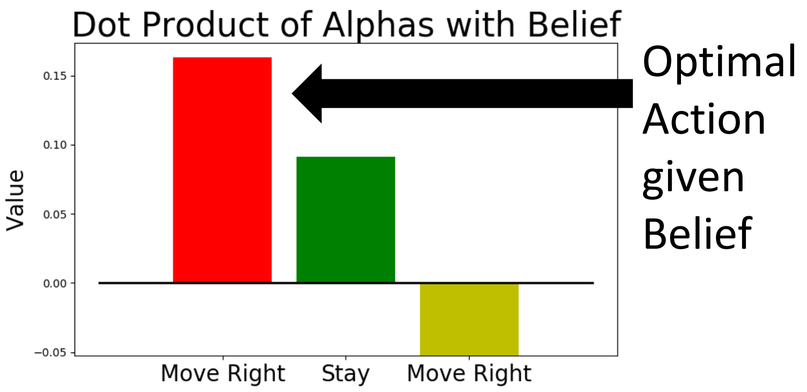
Gaussian Mixture (GM)

$$\sum_j^J w_j \phi(s|\mu_j, \Sigma_j)$$

Can represent arbitrary policy functions & pdfs



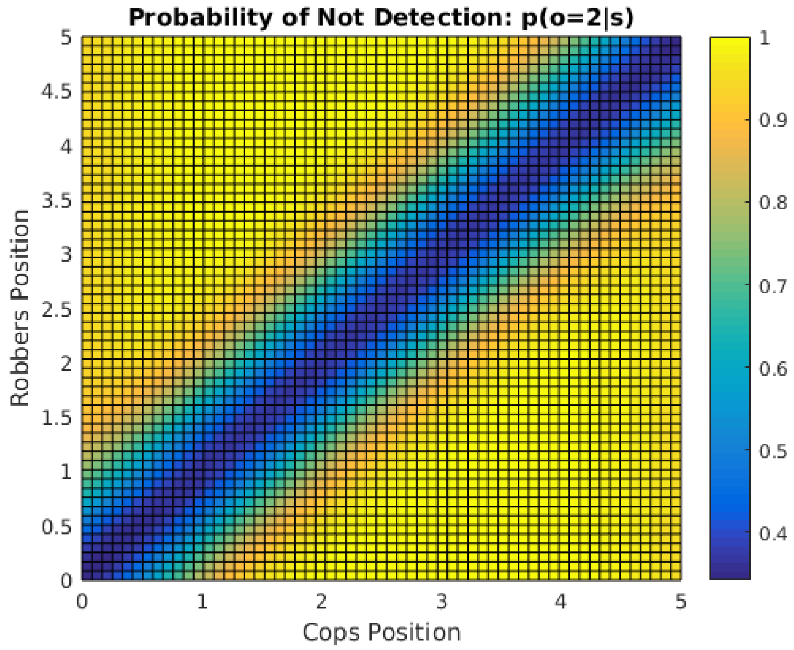




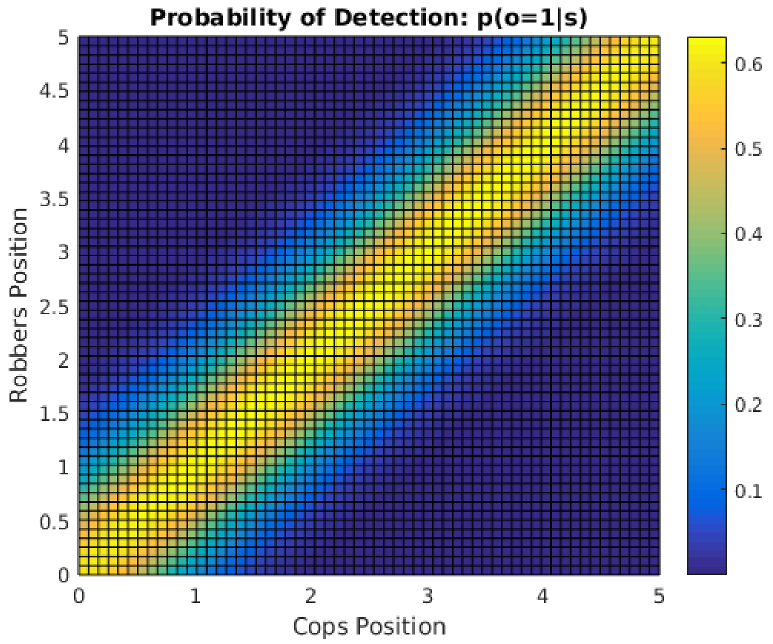
Semantic Observational Models

- Unnormalized GM observation models lead to closed form policy approximations, but non-trivial to specify
- Softmax Models require significantly fewer parameters to construct realistic observation likelihoods than unnormalized GM models
  - Linear scaling with dimension

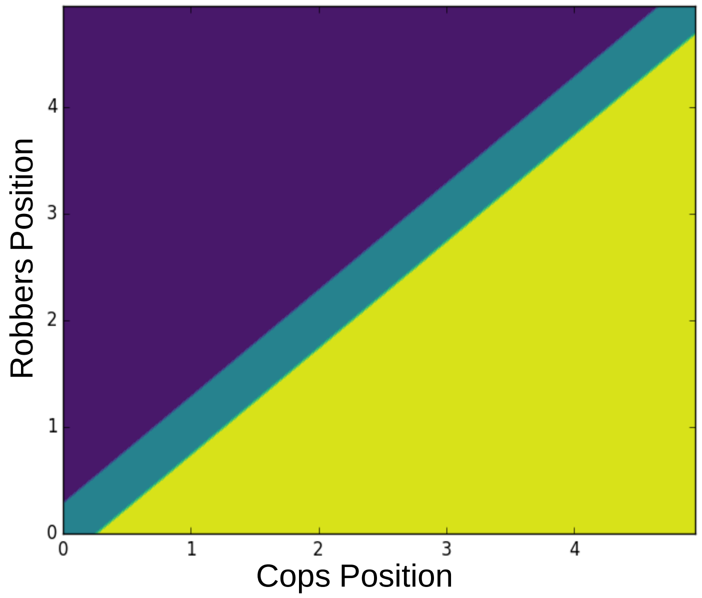
GM with 200 Mixands
600 Parameters



GM with 8 Mixands
24 Parameters



Softmax Model with 3 Classes
9 Parameters



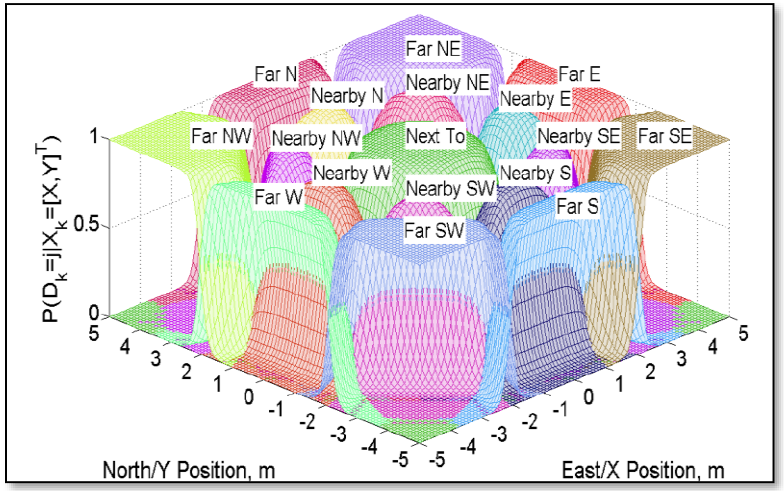
Semantic Observational Models [Sweet, Ahmed, ACC 2016]

A Useful Approach: Softmax Models

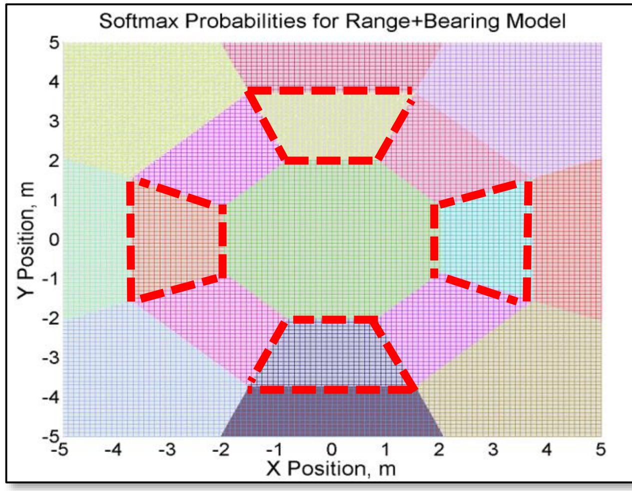
$$p(o_k|s_k) = \frac{\exp(w_o^T s_k + b_o)}{\sum_c^{|\Omega|} \exp(w_c^T s_k + b_c)}$$

- Segment continuous state space into discrete classes
- Classes dominate spatial regions
- Generalizes to non-convex regions
- Sparse parameterization, easy to learn from data and embed constraints

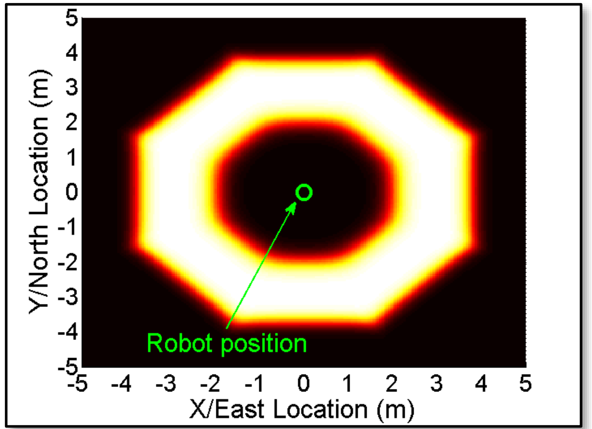
Likelihoods for all classes



Dominant Regions



Compound "Near" Observation Likelihood



One Small Problem

Alpha Element Backup

$$\alpha^i_{a,o}(s) = \int_{s'} \alpha^i_{n-1}(s')p(o|s')p(s'|s,a)ds',$$

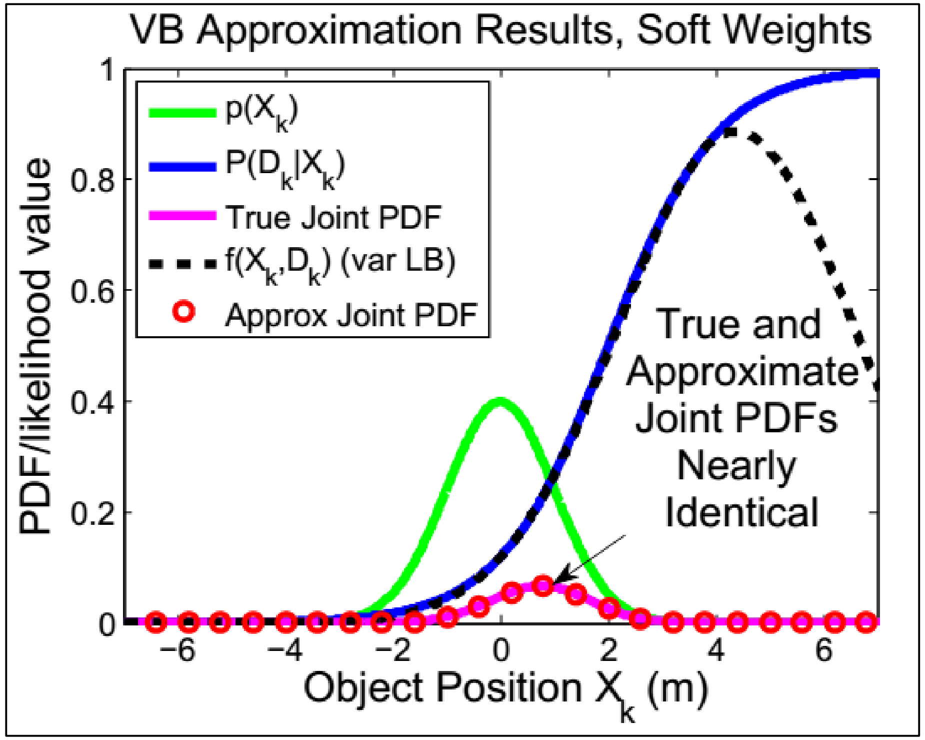
$$\downarrow$$

$$\alpha^i_{n-1}(s')p(o|s') = \sum_j^J \underbrace{w_j |2\pi\Sigma_j|^{-\frac{1}{2}} \exp(-\frac{1}{2}(s'-\mu_j)'\Sigma_j^{-1}(s'-\mu_j)) \frac{\exp(w_o^T s' + b_o)}{\sum_c^{|\Omega|} \exp(w_c^T s' + b_c)}}_{\text{Irreducible!}}$$

- Rather than haul around softmax terms through each successive backup, what if we could approximate the products as GMs?

Variational Bayes (VB) Based on [Ahmed, et al IEEE T-RO, 2013]

- Uses EM to approximate products of softmax and Gaussians as Gaussians
- Adapted to unnormalized alpha functions



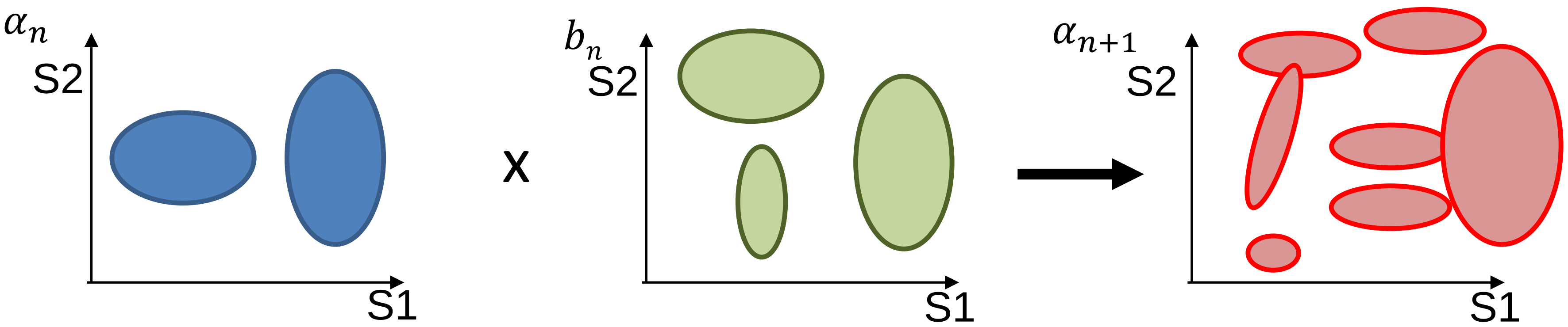
$$\alpha_{n-1}^{i}(s')\; p(o|s') = \left[\sum_k w_k^i \phi(s'|s_k^i, \Sigma_k^i)\right]\left[\frac{\exp w_o^T s' + b_o}{\sum_{c=1}^{S} \exp w_c^T s' + b_c}\right]$$

$$\approx \sum_{h=1}^{H} w_h \phi(s'|\mu_h, \Sigma_h)$$

→ Closed form approximate Bellman backups!

GM Condensation

GM size grows quickly under PBVI backups and belief updates



Alpha Element Backup

$$\alpha^i_{a,o}(s) = \int_{s'} \alpha^i_{n-1}(s')p(o|s')p(s'|s,a)ds'$$

$$\alpha^i_n(s) = r_a(s) + \gamma \sum_o \arg\max_{\alpha^i_{a,o}}(<\alpha^i_{a,o}, b>)$$

If GM observations:

$$|p(o|s')| = 10$$

$$|\alpha_{n-1}(s')| = 10$$

$$|p(s'|s,a)| = 1$$

$$|\boldsymbol{\Omega}| = \mathbf{3}$$

$$|r_a(s)| = 10$$

Then:

$$|\alpha_n(s)| = 310$$

If Softmax observations:

$$|p(o|s')| = 1$$

$$|\alpha_{n-1}(s')| = 10$$

$$|p(s'|s,a)| = 1$$

$$|\boldsymbol{\Omega}| = \mathbf{3}$$

$$|r_a(s)| = 10$$

Then:

$$|\alpha_n(s)| = 40$$

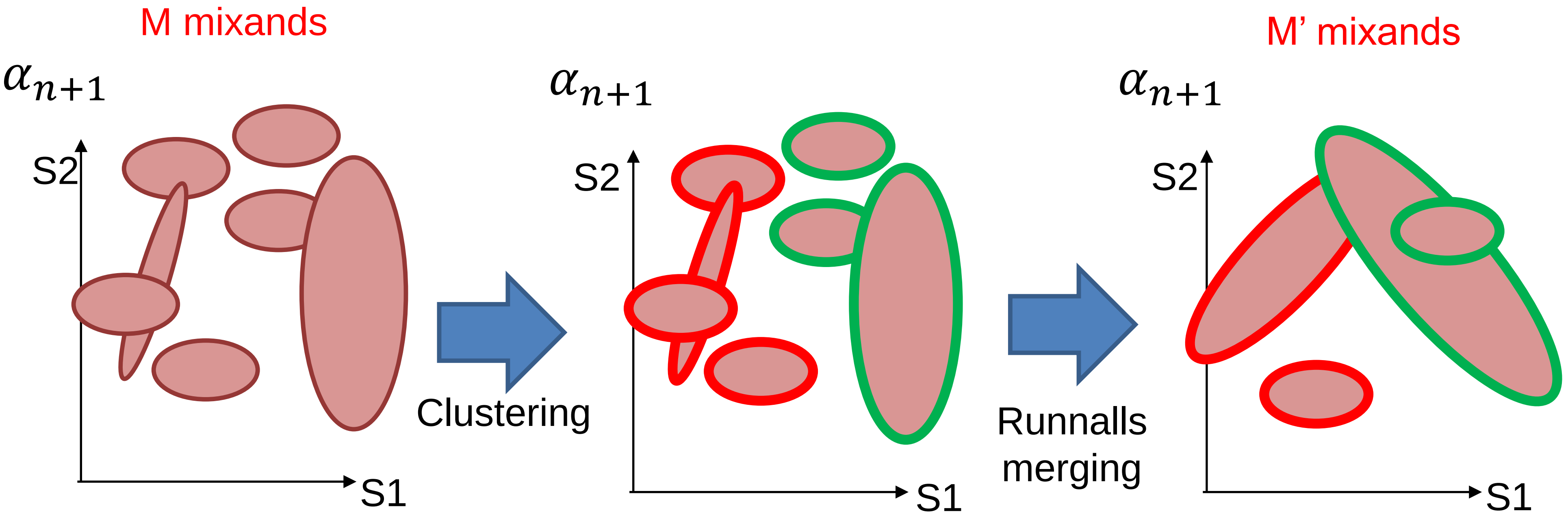
GM Condensation

A method is needed to condense the mixture such that:

$$a_{n+1}(s) = \sum_{k}^{M} w_k \phi(s'|\mu_k, \Sigma_k) \longrightarrow \hat{a}_{n+1}(s) = \sum_{k}^{M'} \hat{w}_k \phi(s'|\hat{\mu}_k, \hat{\Sigma}_k)$$

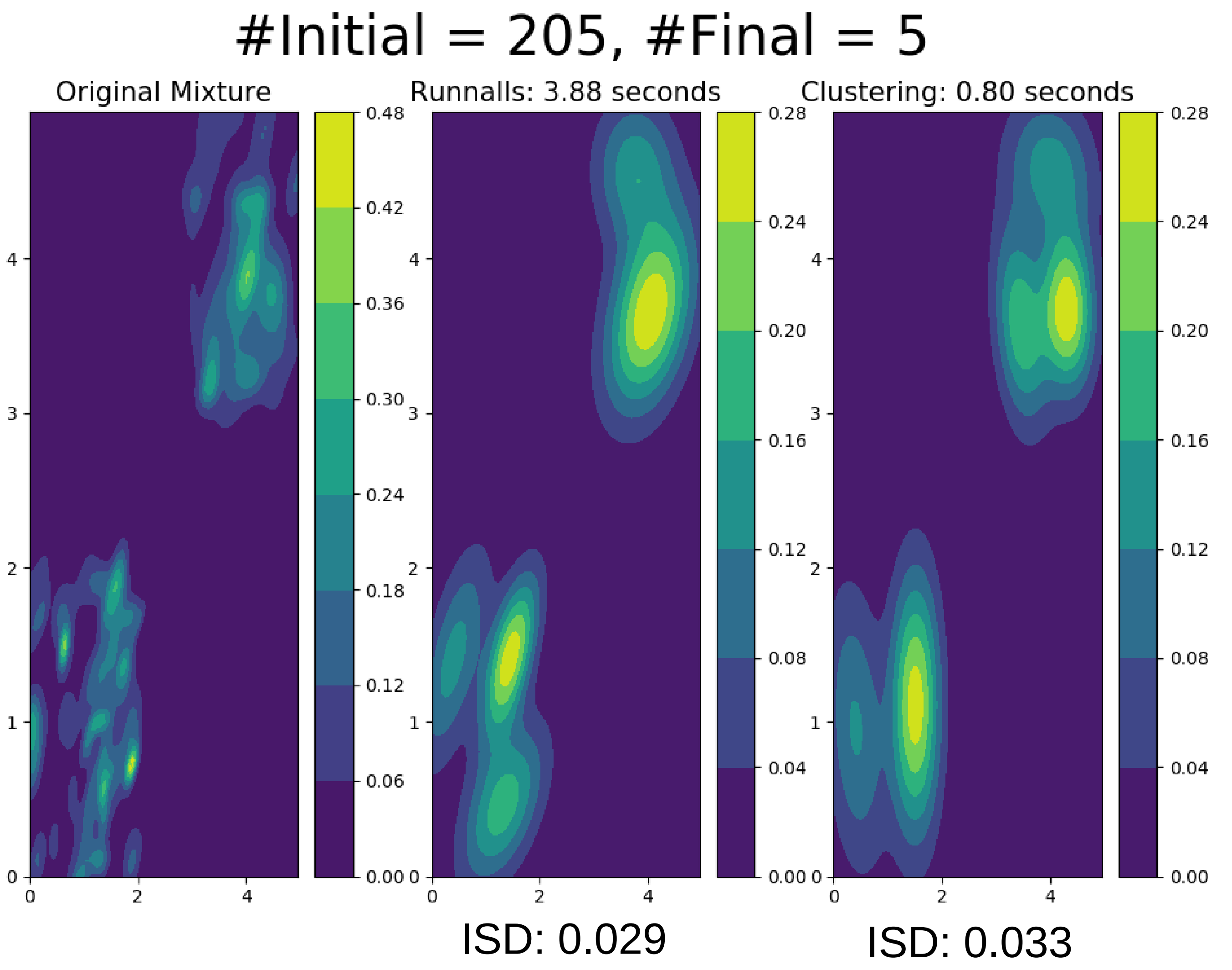
$$\alpha_{n+1}(s) \approx \hat{\alpha}_{n+1}(s) \quad (NK = M' < M)$$

- K-means partitions mixands: μ-Euclidean distance
- [Runnalls, AES 2007] condenses each cluster to N mixands
- Clusters then recombined



Condensation Example

- Heuristically clusters mixands by Euculidean distance of means
- Error measure with the Integral Squared Difference metric (ISD)
  - [Williams, Maybeck, ICIF 2003]
- Parallelizes condensation
- Can tune based on need for accuracy vs. speed
- Continuing work: By what metric should we cluster mixands?



2D Simulation

Higher-dimension Problem:
Cop maintains contact with a Robber in constrained 2D space

States: Defined over difference of dimensions

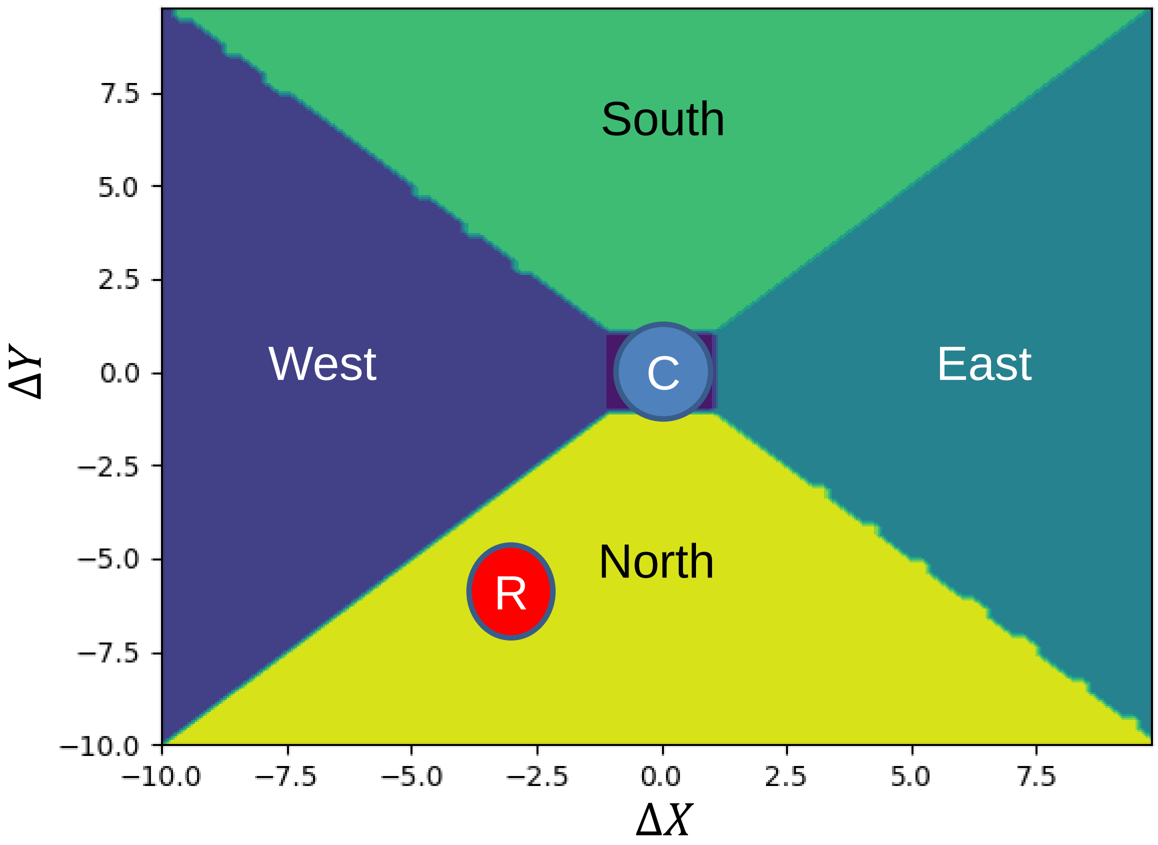
Dynamics:
Robber executes a random walk,
Cop moves in cardinal directions

$$S = [\Delta X, \Delta Y] = [C_x - R_x, C_y - R_y]$$
$$C_x \in (0,5), C_y \in (0,5), R_x \in (0,5), R_y \in (0,5)$$

Observations:
$$\mathbf{\Omega} = \{\mathbf{North}, \mathbf{South}, \mathbf{East}, \mathbf{West}, \mathbf{Near}\}$$

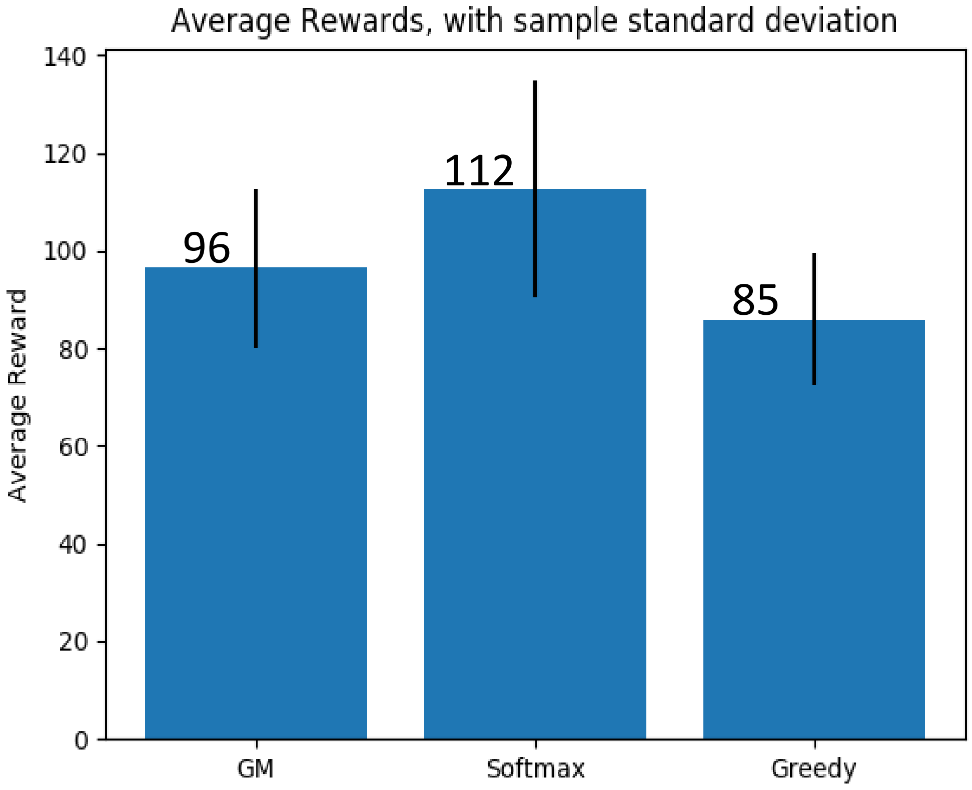
Rewards:
$$r(Dist(R,C) \leq 1) = 3$$
$$r(Dist(R,C) > 1) = 0$$

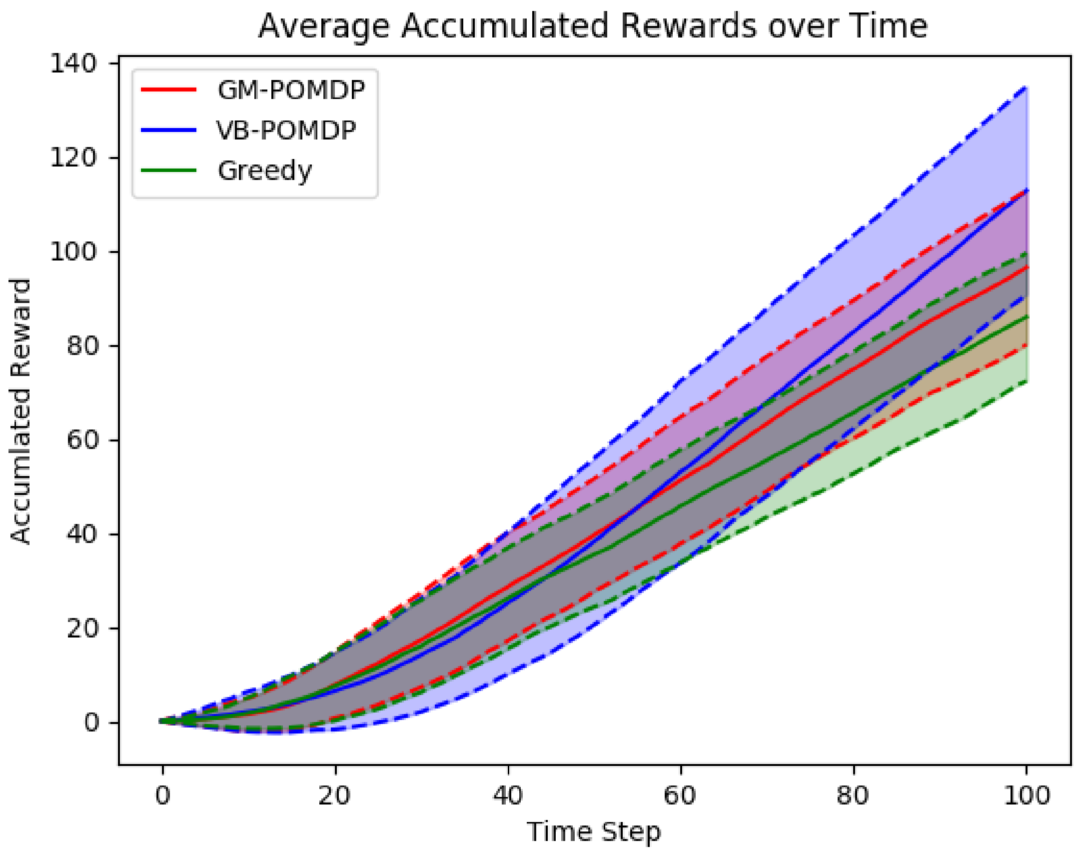
Differencing Observation Model



Comparison of Results (2D)

- Equal computation time given to both the VB-POMDP (9 parameters) and GM-POMDP (~600 parameters)
- VB-POMDP accomplished more backups in that time

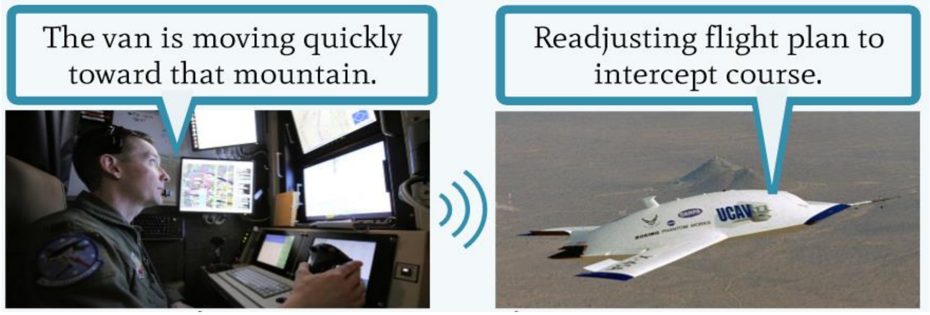


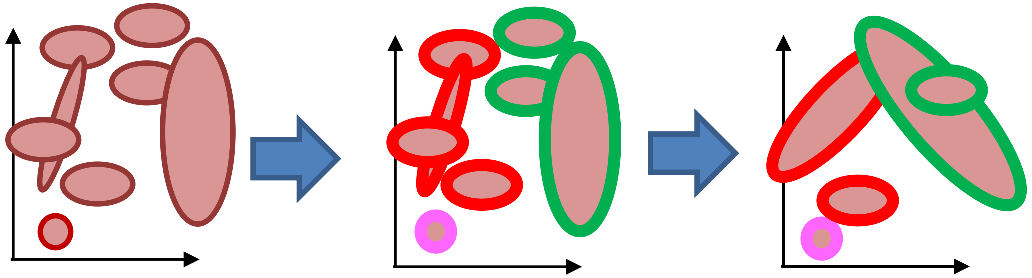


- VB-POMDP outperforms both GM-POMDP and Greedy approach in pairwise comparisons ($p<0.01$)

Conclusions

- Closed form CPOMDP policy function approximations via softmax semantic observation models
  - Insensitive to state space extent/dimension (no discretization)
  - Depends on belief complexity (i.e. # of mixands)

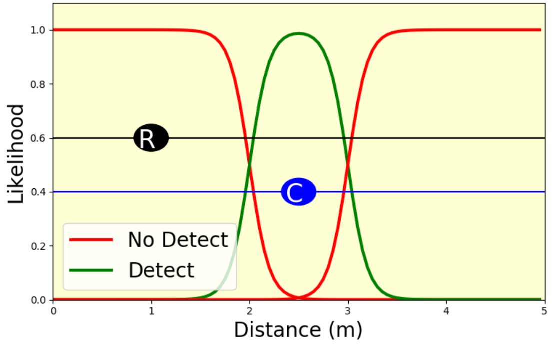




- Heuristic clustering efficiently parallelizes and expedites GM Condensation
  - No significant performance loss

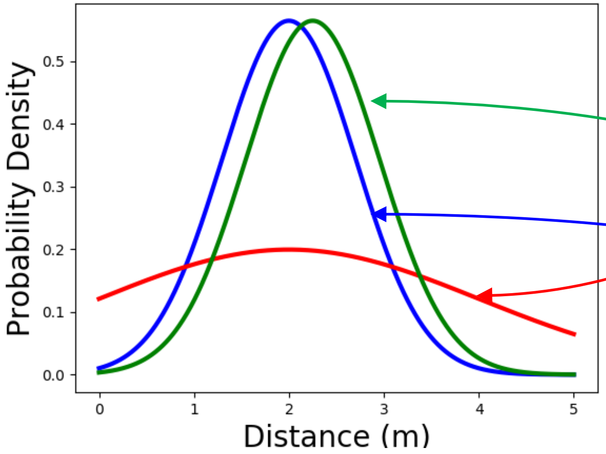
- VB-POMDP improves on state of the art
  - Better scaling for large observation spaces
  - More efficient policy updates → Faster convergence

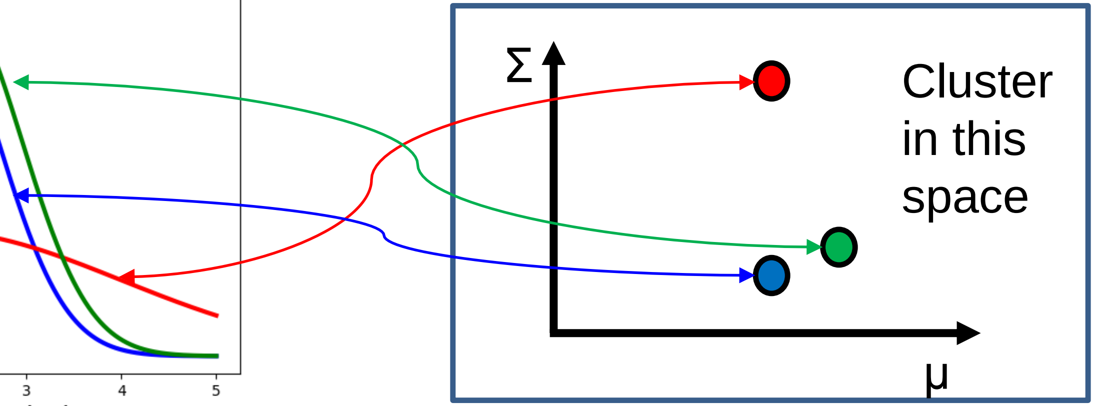
Colinear Detection Zone

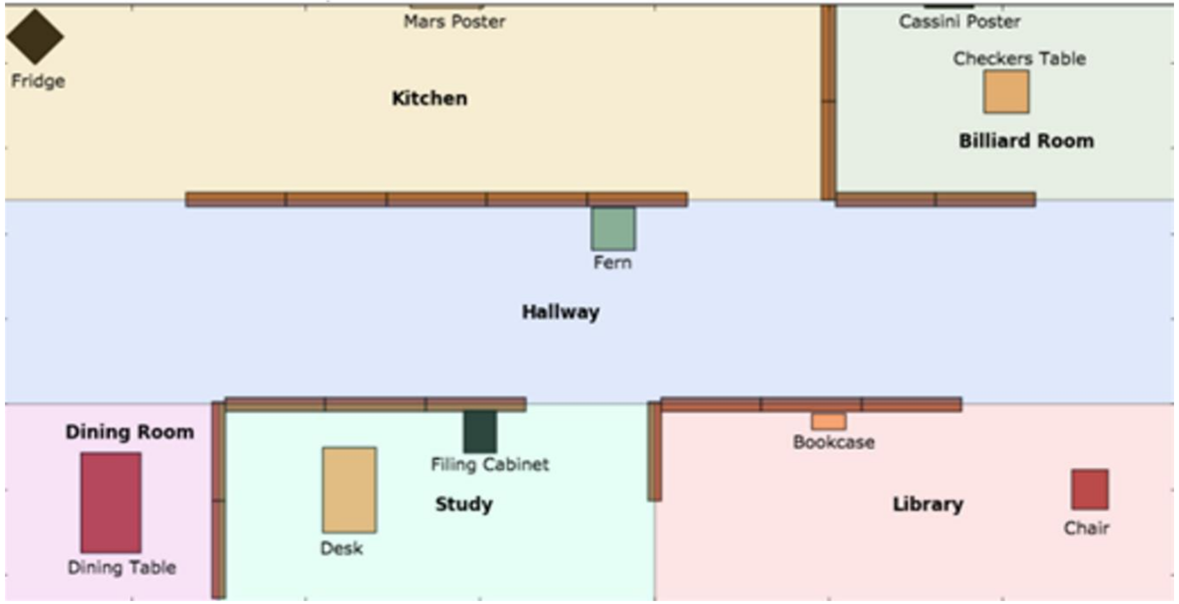


Future Work

- Improve GM Condensation
  - Clustering in PDF space
- Hierarchical Approach
  - Exploit problem structure to scale to more complex problems







- Verification on physical hardware
  - Vision-based target detection and tracking
  - Humans sensors providing semantic observations
  - Natural Language Interface

Research Support Acknowledgement

C-UAS A National Science Foundation Industry/University Cooperative Research Center (IUCRC)