# Bootstrap Learning for Visual Perception on Mobile Robots

## ICRA-11 Uncertainty in Automation Workshop

Mohan Sridharan

Stochastic Estimation and Autonomous Robotics (SEAR) Lab
Department of Computer Science
Texas Tech University

May 9, 2011

# Collaborators

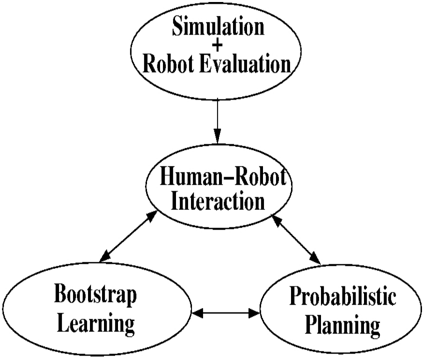
- Mohan Sridharan, Texas Tech University.
- Xiang Li, Shiqi Zhang, Mamatha Aerolla (Graduate Students); Texas Tech University.
- Peter Stone; The University of Texas at Austin.
- Ian Fasel; The University of Arizona.
- Jeremy Wyatt, Richard Dearden; University of Birmingham (UK).

# Desiderata + Challenges

- **Focus:** Integrated systems, visual inputs.
- **Desiderata:**
  - Real-world robots systems require *high reliability*.
  - Dynamic response requires *real-time operation*.
  - Learn from *limited feedback* and operate *autonomously*.
- **Challenges:**
  - *Partial observability:* varying levels of uncertainty.
  - *Constrained processing:* large amounts of raw data.
  - *Limited human attention:* consider high-level feedback.

# Research Thrusts

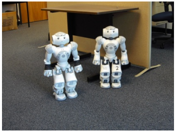
- Learn models of the world and revise learned models over time (*bootstrap learning*).
- Tailor learning and processing to the task at hand (*probabilistic planning*).



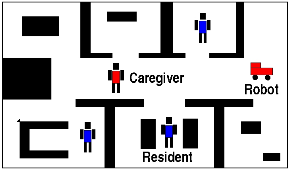

- Enable human-robot interaction with high-level input (*Human-robot Interaction*).

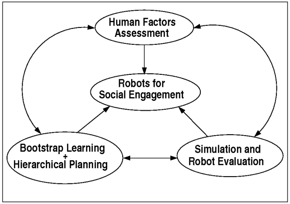
# Robot Platforms and Generalization

- Evaluation on robot platforms and in simulated domains.


- Social engagement in elderly care homes.



# Talk Outline

- Unsupervised learning of object models:
  - Local, global and temporal visual cues to learn probabilistic layered object models.
- Hierarchical planning for visual learning and collaboration:
  - Constrained convolutional policies and belief propagation in hierarchical POMDPs.
- Summary.

# Talk Outline

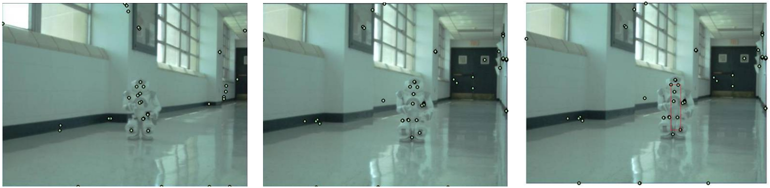
- Unsupervised learning of object models:
  - Local, global and temporal visual cues to learn probabilistic layered object models.
- Hierarchical planning for visual learning and collaboration:
  - Constrained convolutional policies and belief propagation in hierarchical POMDPs.
- Summary.

# Motivation

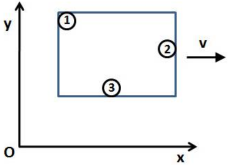
- Learning object models autonomously:
  - Motivation: novel “objects” can be introduced; existing objects can move.
  - Observations: moving objects are interesting! Objects have considerable structure.
- Approach:
  - Analyze image regions corresponding to moving objects.
  - Extract visual features to learn probabilistic object models.
  - Revise models over time to account for changes.

# Tracking Gradient Features

- Tracking and cluster gradient features based on velocity.



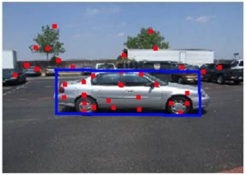
- Model spatial coherence of gradient features.

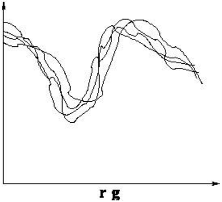


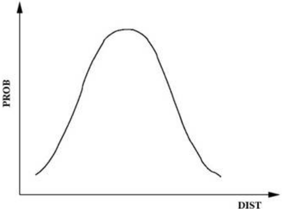

| X | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | -1 | -1 |
| 2 | 1 | 0 | 1 |
| 3 | 1 | -1 | 0 |

| Y | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 1 | 1 |
| 2 | -1 | 0 | 1 |
| 3 | -1 | -1 | 0 |

# Learning Color Features





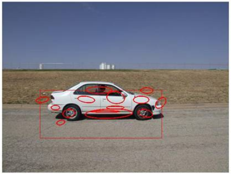



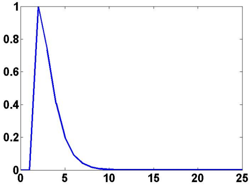

- Use *perceptually-motivated* color space.
- Learn color distribution statistics.
- Learn second-order distribution statistics:

$$JS(a,b) = \frac{1}{2}\{KL(a,m)+KL(b,m)\}, \quad m=\frac{1}{2}(a+b)$$

$$KL(a,m) = \sum_i \{a_i ln \frac{a_i}{m_i}\}$$

# Parts-based Models

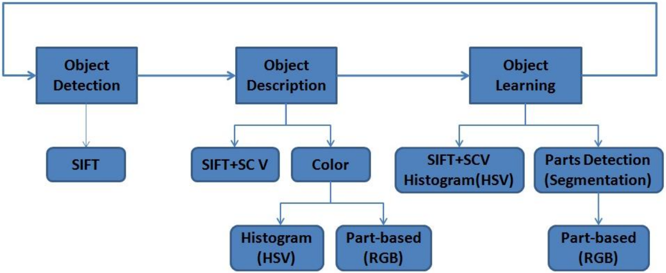





- Graph-based segmentation of input images.
- Gaussian models for individual parts.
- Gamma distribution for inter-part dissimilarity and intra-part similarity.

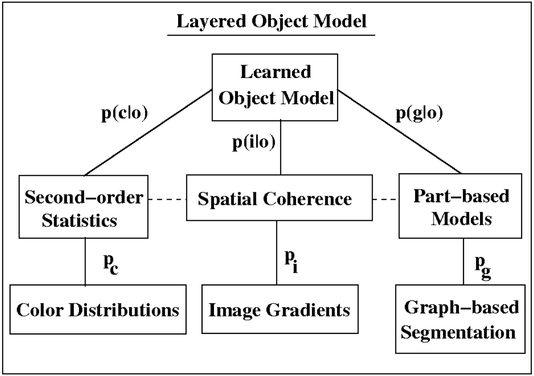
# Layered Object Model

- Model Overview:



# Layered Object Model

- Bayesian belief propagation:



# Recognition

- Stationary and moving objects – motion required only to learn object models.
- Extract features and compare with learned models.
- Find region of relevance based on gradient features.

HUMAN?

CAR?

BOX?

OTHERS?

# Recognition - Gradients

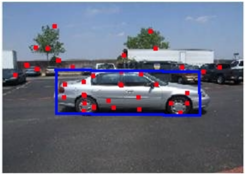
Find probabilistic match using spatial similarity measure.

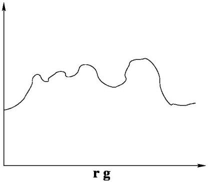
| X | 1 | 2 | ⋯ | N |
|---|---|---|---|---|
| 1 | 0 | -1 | ⋯ | -1 |
| 2 | 1 | 0 | ⋯ | 1 |
| ⋮ | ⋮ | ⋮ | ⋱ | ⋮ |
| N | 1 | -1 | ⋯ | 0 |

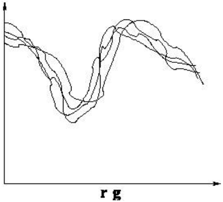
| Y | 1 | 2 | ⋯ | N |
|---|---|---|---|---|
| 1 | 0 | 1 | ⋯ | 1 |
| 2 | -1 | 0 | ⋯ | 1 |
| ⋮ | ⋮ | ⋮ | ⋱ | ⋮ |
| N | -1 | -1 | ⋯ | 0 |

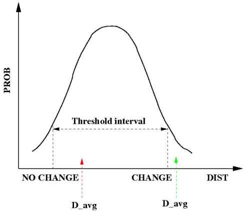
$$SSM(scv_i, scv_{test}) = \frac{N^{i,test}_{x,correct} + N^{i,test}_{y,correct}}{2(N-1)}, \quad SSM \in [0, 1]$$

# Recognition - Color Distributions






HUMAN?

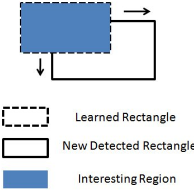
CAR?

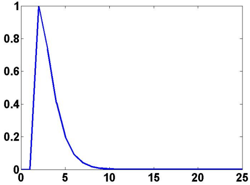
BOX?

OTHERS?

# Recognition - Parts-based Models

- Dynamic programming to match learned models over the relevant region.



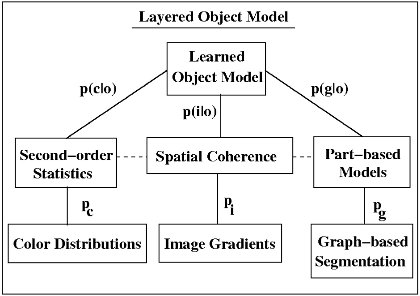

- Similarity within a part, dissimilarity between parts.

$$p_j^{i,arr} = f(sim) \cdot f(diff)$$

$$p^{i,arr} = \sum_j w_j^{l_i} \cdot p_j^{i,arr}$$

# Recognition - Overall

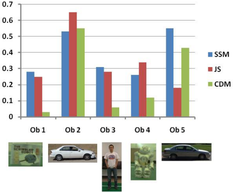
- Combine evidence from individual visual features.
- Bayesian update for belief propagation.

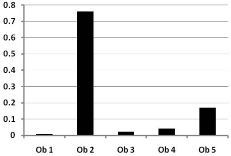



- Recognize objects or identify novel objects.

# Experimental Results

Good classification and recognition performance.




| $p(O\|A)$ | Box | Human | Robot | Car | Other |
|---|---|---|---|---|---|
| Box | 0.913 | 0.013 | 0.02 | 0 | 0.054 |
| Human | 0.027 | 0.74 | 0.007 | 0.013 | 0.213 |
| Robot | 0.033 | 0.007 | 0.893 | 0 | 0.067 |
| Car | 0 | 0.02 | 0 | 0.833 | 0.147 |

# Talk Outline

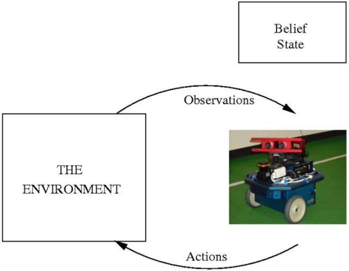
- Unsupervised learning of object models:
  - Local, global and temporal visual cues to learn models.
- Hierarchical planning for visual learning and collaboration:
  - Constrained convolutional policies and belief propagation in POMDPs.
- Summary.

# Motivation

- Large amount of data, many processing algorithms.
- Cannot learn all models comprising all possible features!
- Sensing and processing can vary with task and environment:
  - Where do I look? What do I look for?
  - How to process the data?
- Approach: tailor sensing and processing to the task.
  - Partially Observable Markov Decision Processes (POMDPs).

# POMDP Overview

- Tuple: $\langle \mathcal{S}, \mathcal{A}, \mathcal{Z}, \mathcal{T}, \mathcal{O}, \mathcal{R} \rangle$
- Belief distribution $B_t$ over states.
- Actions $\mathcal{A}$.
- Observations $\mathcal{Z}$: action outcomes.
- Transition function: $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S}' \mapsto [0, 1]$




- Observation function $\mathcal{O} : \mathcal{S} \times \mathcal{A} \times \mathcal{Z} \mapsto [0, 1]$
- Reward specification $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \Re$
- Policy $\pi : B_t \mapsto a_{t+1}$

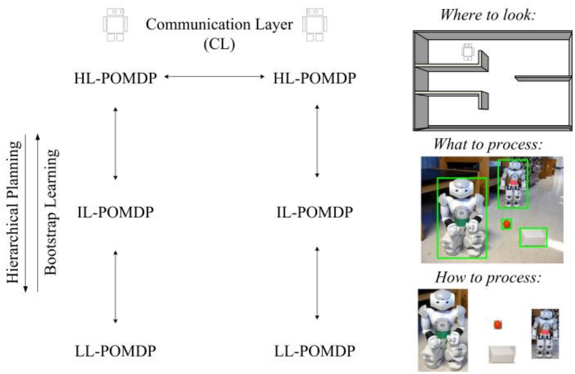
# Challenges

- State space increases exponentially.
- Policy generation methods are exponential (worst-case) in the state space dimensions.

- Model definition may not be known and may change.
- Intractable for real-world applications!

- Observations:
  - Only a subset of scenes and inputs are relevant to any task.
  - Visual sensing and processing can be organized hierarchically.

# Hierarchical Visual Planning



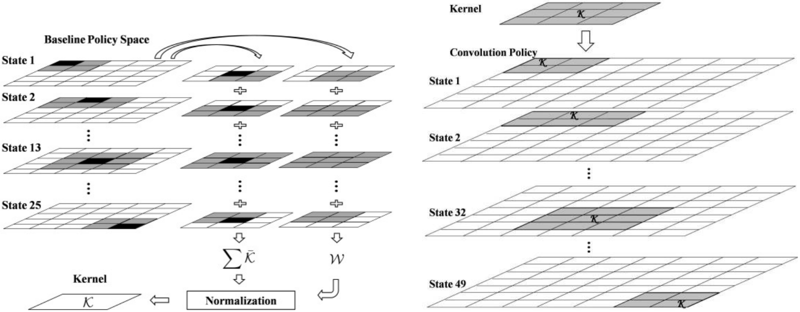

- Constrained convolutional policies.
- Automatic belief propagation.

# HL Search – Convolutional Policies

- Rotation and shift invariance of local visual search.

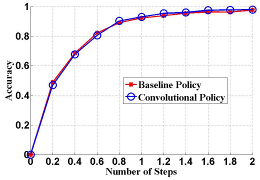
$$\bar{K}(s) = (\pi^H \otimes C_m^K)(s) = \int \pi^H(\tilde{s}) C_m^K(s - \tilde{s}) d\tilde{s}, \quad K = (\sum_{a_i} \bar{K}) \cdot / W$$

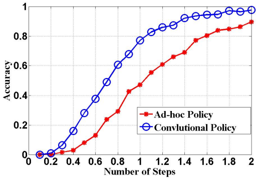
$$\pi_C^H(s) = (K \otimes C_m^E)(s) = \int K(\tilde{s}) C_m^E(s - \tilde{s}) d\tilde{s}$$



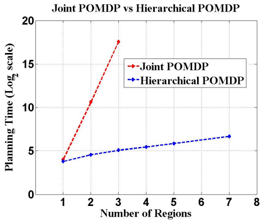
# Experimental Results

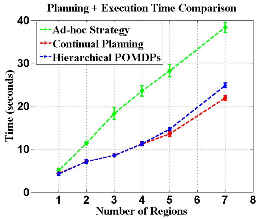
- Accurate and efficient visual search.



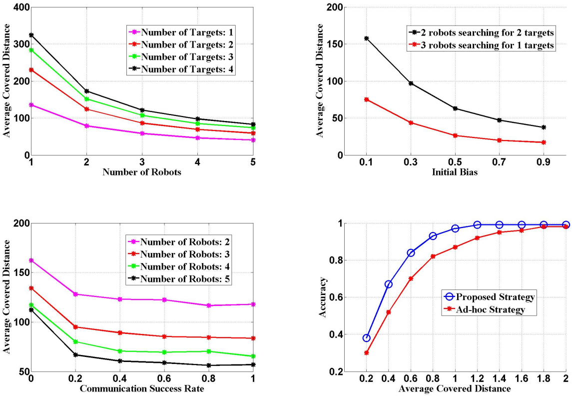

- Reliable (93% vs 87%) and autonomous processing.



# Multirobot Collaboration

Extension to multirobot collaboration (96% vs. 88%).



# Talk Outline

- Unsupervised learning of object models:
  - Local, global and temporal visual cues to learn models.
- Hierarchical planning for visual learning and collaboration:
  - Constrained convolutional policies and belief propagation in POMDPs.
- Summary.

# Summary

- Robot autonomously acquires models for different object categories. Detects and tracks objects in subsequent images with high ($\geq 90\%$) accuracy.
- Hierarchical planning enables a team of robots to share beliefs and collaborate robustly in dynamic domains.
- Learning and hierarchical planning inform and guide each other to result in autonomous (and real-time) operation of mobile robots in complex environments.

# Additional Challenges

- Learn correlations between visual cues to learn better object models.
- Assess quality of (information in) object models. Infer lack of information and the presence of novel objects.
- Reason with non-visual inputs by incorporating hierarchical decompositions that match corresponding cognitive requirements.

# Recent Papers I


- Xiang Li, Mohan Sridharan and Shiqi Zhang. **Autonomous Learning of Vision-based Layered Object Models on Mobile Robots.** To Appear In the International Conference on Robotics and Automation (ICRA 2011), Shanghai, China, May 9-13, 2011.

- Shiqi Zhang, Mohan Sridharan and Xiang Li. **To Look or Not to Look: A Hierarchical Representation for Visual Planning on Mobile Robots.** To Appear In the International Conference on Robotics and Automation (ICRA 2011), Shanghai, China, May 9-13, 2011.

# Recent Papers II


- Xiang Li and Mohan Sridharan. **Safe Navigation on a Mobile Robot using Local and Temporal Visual Cues.** In the International Conference on Intelligent Autonomous Systems (IAS 2010), Ottawa, Canada, August 30-September 1, 2010.
- Mohan Sridharan, Jeremy Wyatt and Richard Dearden. **Planning to See: A Hierarchical Approach to Planning Visual Actions on a Robot using POMDPs.** Artificial Intelligence Journal, Volume 174, Issue 11, pages 704-725, July 2010.

- All papers available for download: `www.cs.ttu.edu/~smohan/Publications.html`

# We are done!

Questions? Comments?