Exploiting Semantics in Mobile Robotics

A. Aydemir, K. Sjöö, A. Pronobis and P. Jensfelt

Computer Vision and Active Perception Lab
Centre for Autonomous Systems
KTH - Royal Institute of Technology
Stockholm, Sweden
Why Semantics in Mobile Robotics?

• More efficient task execution by exploiting semantic information embedded in man-made environments

• Examples:
  - Navigation in unknown environment / goal-driven exploration
  - Fetch and carry (find objects)
  - Location-specific actions
  - Knowledge transfer between human (and human knowledge sources) and robot
  - Human-robot-interaction
Representing and Reasoning about Space From Low Level to High Level

- High-level knowledge
  - Human level concepts (e.g. rooms, obj-obj relations, ...)

- Long-term categorical knowledge
  - Object models, ...

- Discretized space
  - Places, paths, ...

- Low level sensor data
  - Navigation, manipulation, ...
Probabilistic Semantic Mapping

- Detected doors
- Large, elongated corridor
- Place
- Topological graph
- Distribution over room categories and beliefs about property values
- Large, square double office
- Book
Sensing and our models of the world are uncertain
Chain Graph model

Approximate inference: Loopy Belief Propagation

[Pronobis and Jensfelt, ICRA 2012]
Result: Beliefs
Modeling Object Location
• Using functional spatial relations (IN and ON) to model object-object relations and object-location relations.
  - The apples are IN the bowl ON the table
  → Hierarchical decomposition
  → Appropriate level of abstraction for exploiting common sense human knowledge
• A set of axioms for relational predicates $In(x,y), On(x,y)$
We can deduce additional relations:
- “the balls are in the bowl” + “the bowl is on the table” → “the balls are on the table”
- Indirect search: look for a table to find the balls
Results: Scene Analysis

- Axioms + perceptual models + chain graphs

(a) Example 1: “A on B on C”

<table>
<thead>
<tr>
<th></th>
<th>Example 1</th>
<th>Inf</th>
</tr>
</thead>
<tbody>
<tr>
<td>$On(A,B)$</td>
<td>92.5%</td>
<td>TRUE</td>
</tr>
<tr>
<td>$On_t(A,B)$</td>
<td></td>
<td>TRUE¹</td>
</tr>
<tr>
<td>$In(A,B)$</td>
<td>0%</td>
<td>FALSE</td>
</tr>
<tr>
<td>$On(A,C)$</td>
<td>4.4%</td>
<td>FALSE</td>
</tr>
<tr>
<td>$On_t(A,C)$</td>
<td></td>
<td>TRUE²</td>
</tr>
<tr>
<td>$In(A,C)$</td>
<td>0%</td>
<td>FALSE</td>
</tr>
<tr>
<td>$On(B,A)$</td>
<td>0%</td>
<td>FALSE</td>
</tr>
<tr>
<td>$On_t(B,A)$</td>
<td></td>
<td>FALSE</td>
</tr>
<tr>
<td>$In(B,A)$</td>
<td>0%</td>
<td>FALSE</td>
</tr>
<tr>
<td>$On(B,C)$</td>
<td>96.4%</td>
<td>TRUE</td>
</tr>
<tr>
<td>$On_t(B,C)$</td>
<td></td>
<td>TRUE¹</td>
</tr>
<tr>
<td>$In(B,C)$</td>
<td>0%</td>
<td>FALSE</td>
</tr>
</tbody>
</table>

[Sjöö, Pronobis and Jensfelt, ICAR 2011]
Using 3D Context

- Location of objects strongly correlated with local 3D structure (depth cues)
- We call this 3D context of an object
- Idea: learn the mapping between the 3D context and object category
- Related: Torralba, model for predicting object locations using global scene context from appearance based features
Results: Predict Potential Object Locations

- Drastic reduction of search space

Cup

Whiteboard Marker
Kinect@Home

- We need data collected in the wild of real-environments
- The Microsoft Kinect sensor offers high quality RGB-D data for an extremely low price
- Is at the homes of 30+ million people, fastest selling consumer device in history
- Make it dead-simply to collect data using only a browser plug-in.

http://kinectathome.com
Modeling Topology and Unexplored Space
Reasoning About the Unknown

• In an unknown environment we are often required to reason about what “could be”
• Evaluate “world hypotheses”
• Can be formulated in the chain graph by looking at potential extensions to the graph
• How to learn the models?

[Pranobis and Jensfelt, ICRA 2012]
Use Lots of Data!

- Using data from MIT and KTH campuses
- 197 buildings, 940 floors and 38,000 rooms
- Each floor includes:
  - Rooms with labels
  - Connectivity of rooms
  - 2D layout

http://www.cas.kth.se/floorplans
Results: Predicting Topology

• Indoor environments strongly structured
  - Consist of functional sub-parts

• Predicting topology:
  - Method 1: "count-based"
  - Method 2: Exploit sub-parts
Segmentation of Space
Semantics and Function

• One room or many different functional regions? (kitchen, dining area, livingroom, etc.)
Functional Approach

- Encode function explicitly
  - Corridor: Connects other regions
  - Kitchen: For preparing food
  - *Definitions from the Oxford online dictionary*
- Formulate segmentation as energy-maximization problem
- Solve using Simulated Annealing
How much does semantics help, really?
Task: Large-scale Object Search
“Find me the cornflakes”
Observations

• The **object is small** and the **environment is large**! → brute force search infeasible

• Exploit:
  - Semantic knowledge
  - Structure of indoor environments
to reduce the search space

• Examples:
  - Topological object relations (cups on tables)
  - Typical relations between objects or locations (cups in kitchens)
  - Typical room connectivity (kitchens connected to corridors)
Approach

• Combine:
  - Semantic mapping
  - Object relations
  - Predictions for unexplored space
  - Planning

• Probabilistic cost-benefit trade-off analysis:
  - exploring unknown space
  - searching for object in the known part
  Also for the unknown part of space!

[Aydemir et al., ECMR'11], [Aydemir, Pronobis and Jensfelt, T-RO (Submitted)]
Results: Comparing to Human Performance

- Humans solving the same object search task
Results: Comparing Strategies

• Quickly determines if room is relevant

• Similar behavior, needs better viewpoint
Conclusions

- Large-scale Semantic Mapping
  - Uses multiple modalities: objects, geometry, appearance, topology, human input
  - Reasoning about room categories, objects, spatial properties
- Abstracting and modeling object-object and object-location relations and 3D context
- Reasoning about unexplored space and modeling topology
- Functional segmentation of space
- Large-scale object search system
- Using semantic knowledge from semantic mapping system results in efficient, intuitive and human-like behavior
THE END

http://www.pronobis.pro    http://www.csc.kth.se/~aydemir
http://www.csc.kth.se/~krsj    http://www.csc.kth.se/~patric