An Integrated Robotic System for Spatial Understanding and Situated Interaction in Indoor Environments

Hendrik Zender[1] and **Patric Jensfelt**[2] and **Óscar Martínez Mozos**[3] and **Geert-Jan M. Kruijff**[1] and **Wolfram Burgard**[3]

[1]Language Technology Lab, German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany

[2]Centre for Autonomous Systems, Royal Institute of Technology (KTH), Stockholm, Sweden

[3]Department of Computer Science, University of Freibug, Freiburg, Germany

`http://www.cognitivesystems.org/explorer.asp`

Abstract

A major challenge in robotics and AI lies in creating robots that are to cooperate with people in human-populated environments, e.g. for domestic assistance or elderly care. Such robots need skills that allow them to interact with the world and the humans living and working there.

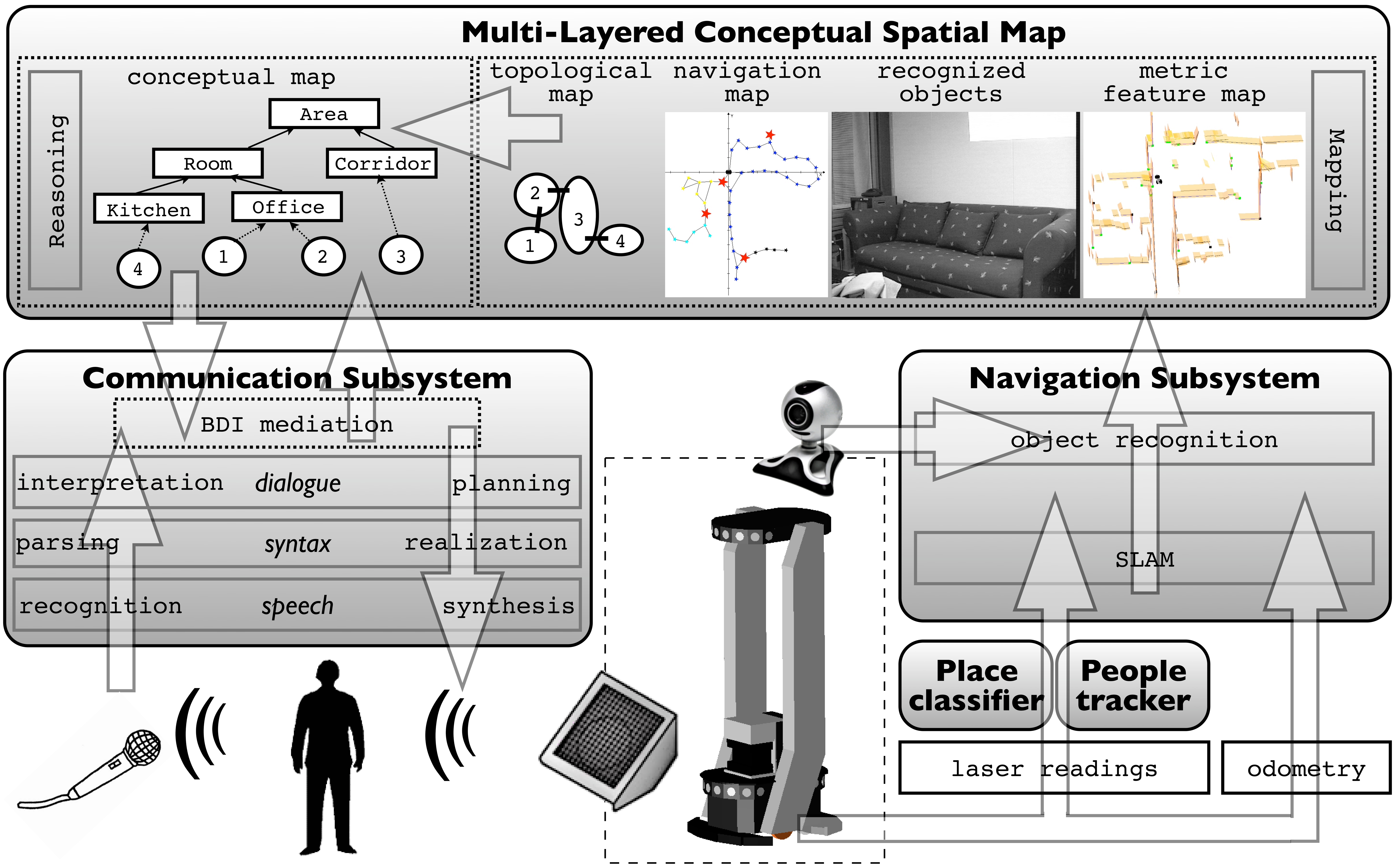
In this work we investigate the question of **spatial understanding of human-made environments**. The functionalities of our system comprise **perception** of the world, **natural language**, **learning**, and **reasoning**. For this purpose we **integrate state-of-the-art components** from different disciplines in AI, robotics, and cognitive systems into a mobile robot system.

Here we describe the principles that were used for the integration, including cross-modal ontology-based mediation, and processing of perception on multiple levels of abstraction. Finally, we present experiments with the **integrated "CoSy Explorer" system** and list some major lessons that were learned from its design, implementation, and evaluation.

Goal

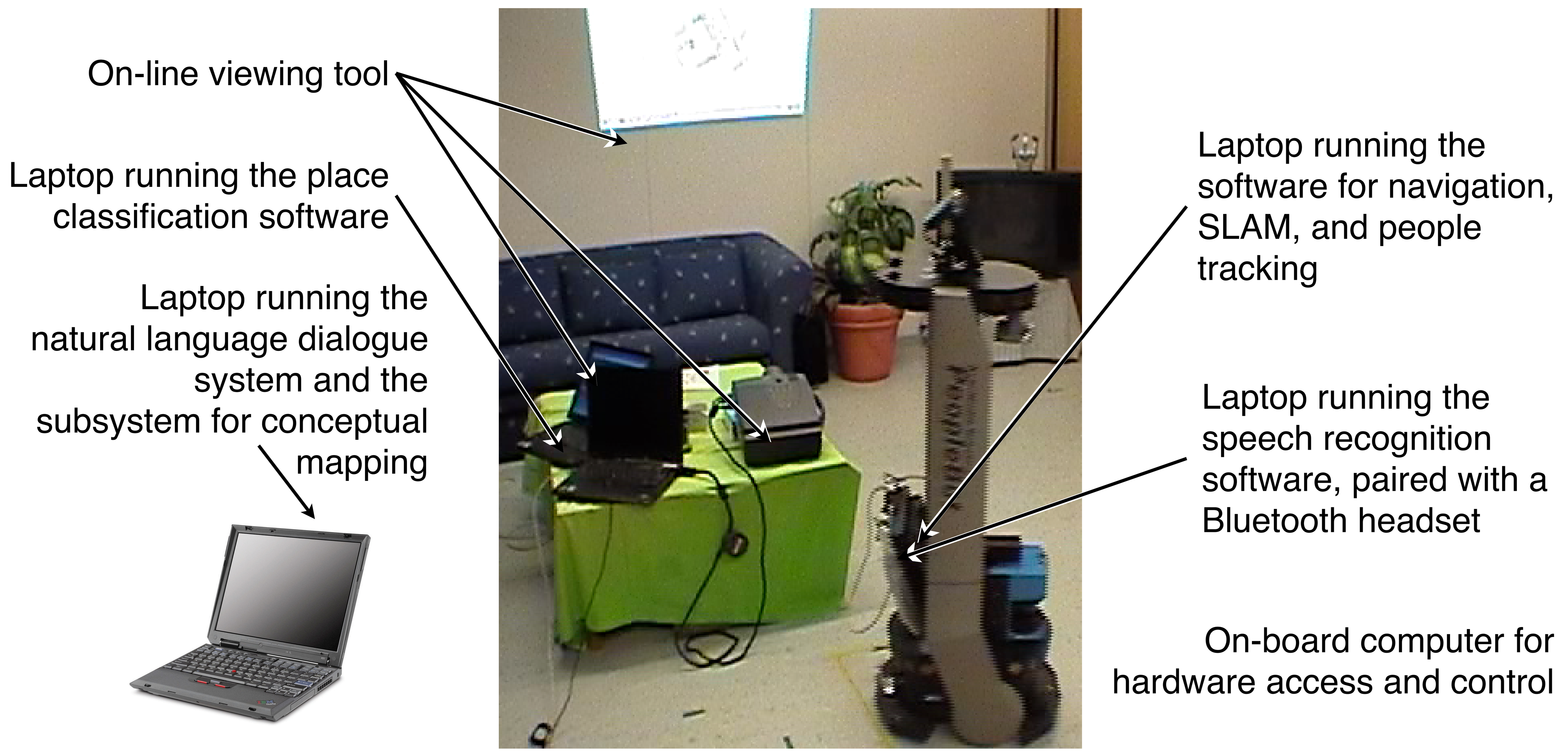
- **Question:** How can a robot understand the semantic, social, and functional aspects of its (human-made) environment?
- **Method:** Create an integrated, cognitive robotic system, using
  - state-of-the-art subsystems
  - cognitive architecture framework
- **Scenario:** *Robot office assistant*
  - interactive, semi-supervised map acquisition (Human-Augmented Mapping, HAM)
  - situated dialogue between robot and user about their environment

The Robot Architecture



The hardware used

Five laptops interconnected via wireless network



The robot platform

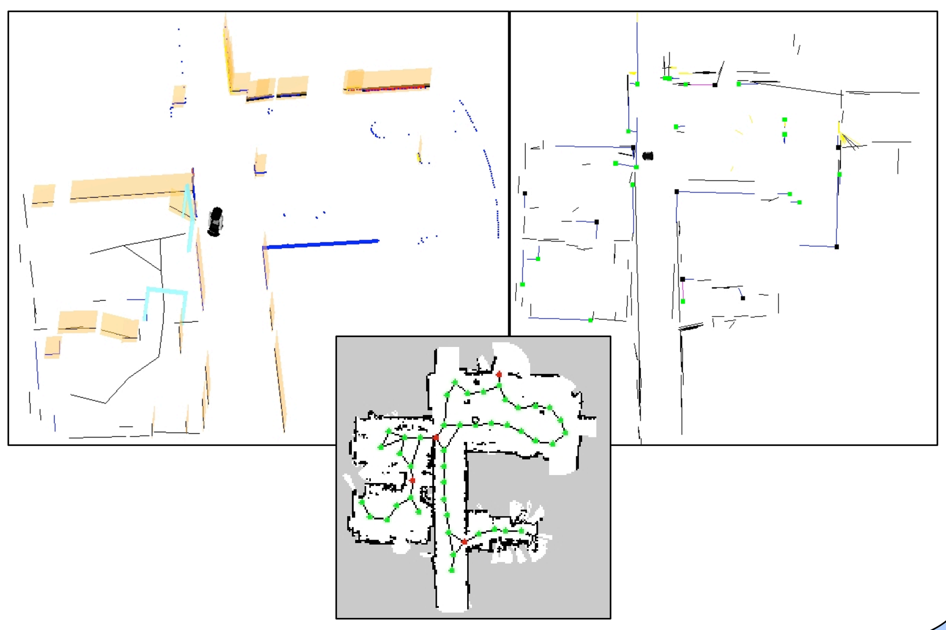
One ActivMedia PeopleBot

System integration, main components, and techniques used

Perception

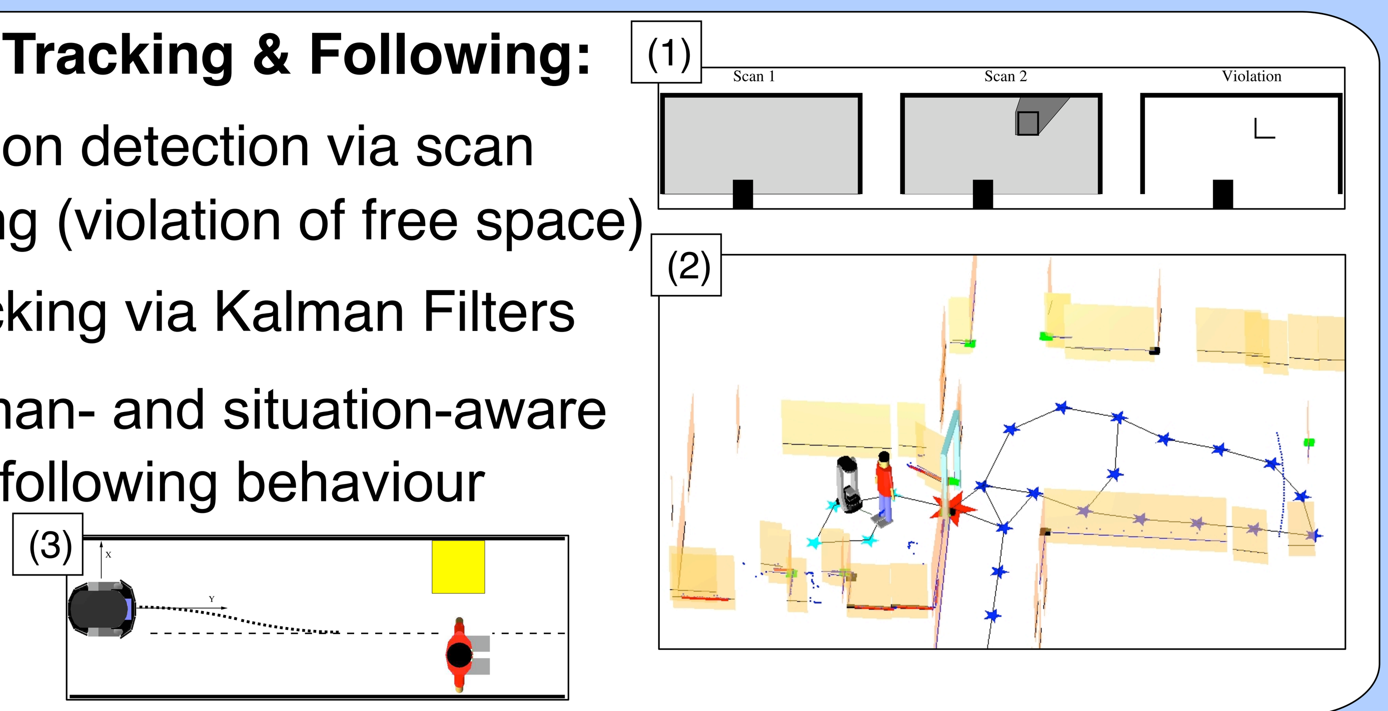
Metric Mapping and Localization:

- Simultaneous Localization and Mapping (SLAM)
- Geometric features (lines) extracted from laser range scans
- Integration of feature measurements using the Extended Kalman Filter (EKF)



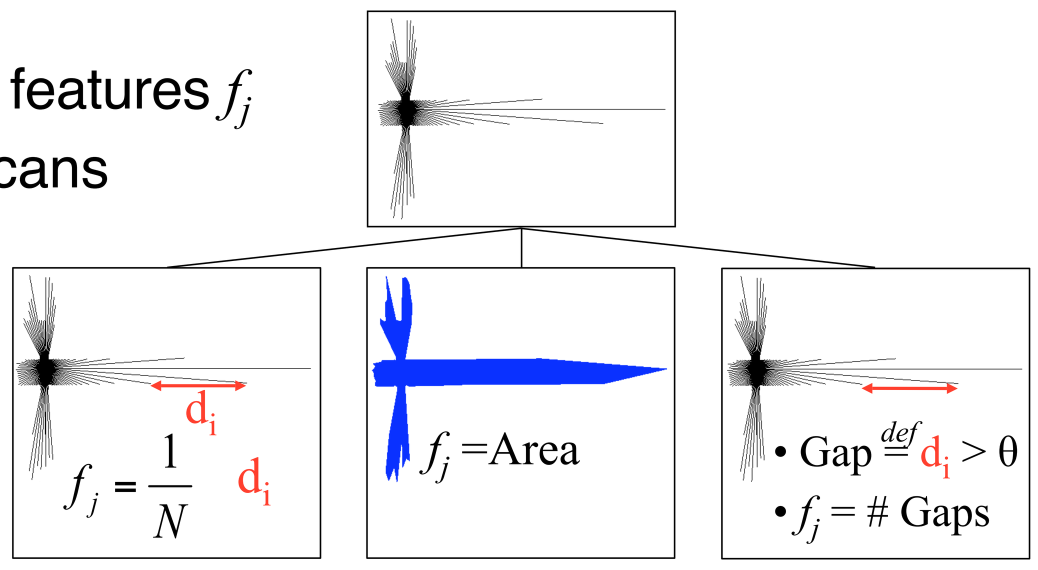
People Tracking & Following:

- (1) Motion detection via scan matching (violation of free space)
- (2) Tracking via Kalman Filters
- (3) Human- and situation-aware person following behaviour



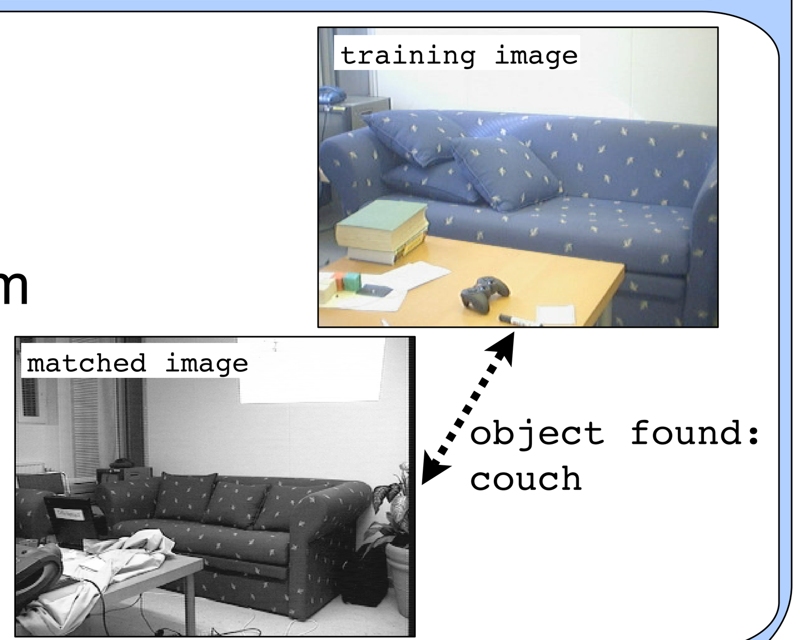
Place Categorization:

- Based on simple geometrical features f_j extracted from laser range scans
- Features are combined using AdaBoost
- Distinguish between *Room* and *Corridor*



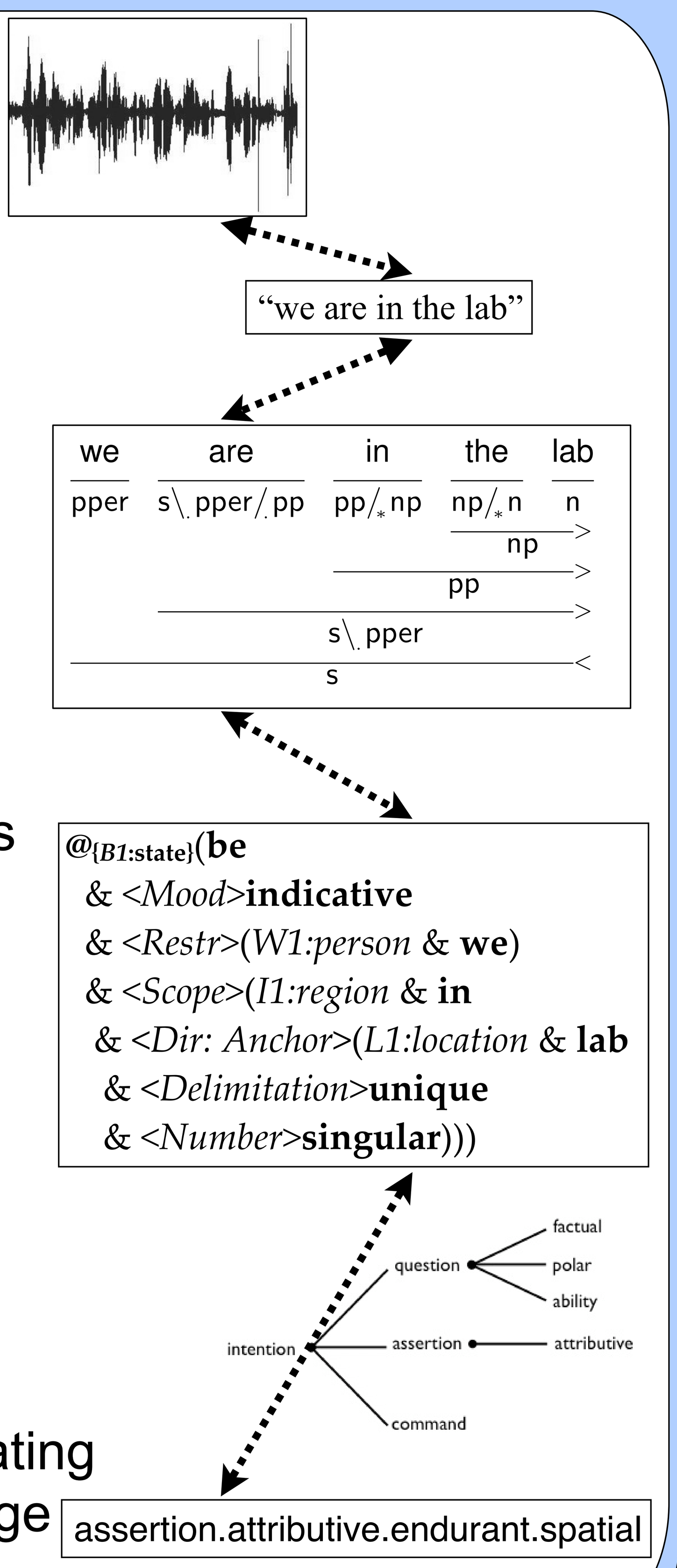
Object Recognition:

- Implementation of the Scale-Invariant Feature Transform (SIFT) computer vision algorithm
- Appearance-based image recognition
- Recognition of instances rather than classes



Language & Dialogue

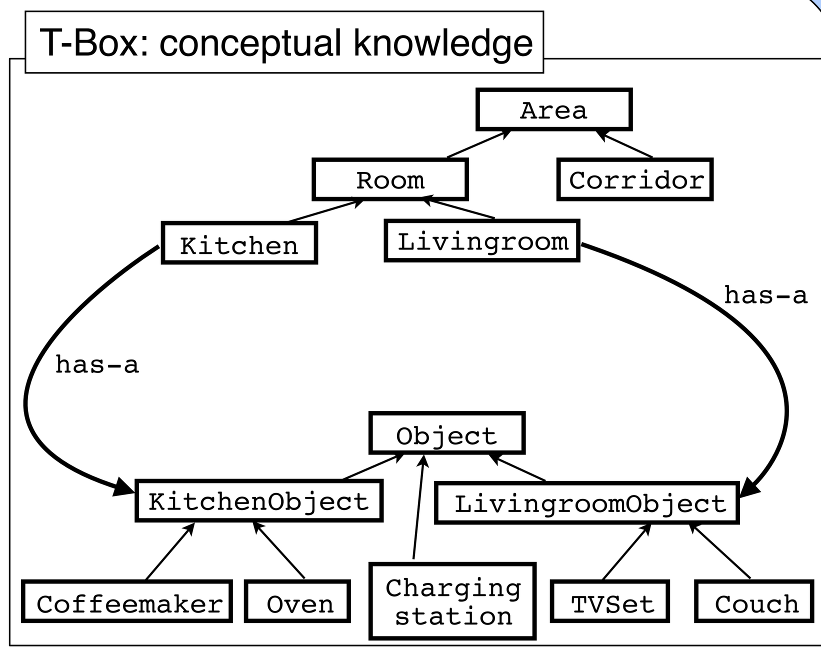
- **Speech Recognition & Synthesis:**
  - Nuance v. 8.5, speaker-independent speech recognition
  - Festival, FreeTTS, or MARY speech synthesis
- **Parsing & Generation:**
  - OpenCCG, combinatory categorial grammar
  - Ontologically rich relational syntactic and semantic representation
- **Semantic Analysis:**
  - Hybrid Logics Dependency Semantics
- **Dialogue Interpretation & Management:**
  - Contextual reference resolution
  - Basic rhetorical relation resolution
  - SDRT-like dialogue context model
  - SFG-like functional interpretation and production of dialogue
- **Cross-modal Information binding:**
  Ontology-based mediation for associating linguistic interpretations with knowledge about the robot's environment



Conceptual Spatial Mapping & Reasoning

- OWL-DL commonsense ontology of an indoor office environment, encoding relations between the different areas and the objects found there
- Description-Logics based A-Box and T-Box reasoning (e.g. RACER or Pellet)
- Combines
  - information from the robot's sensors (laser & vision) (*acquired knowledge*),
  - information given by the robot's tutor (*asserted knowledge*),
  - and conceptual knowledge (*innate*)
  - in order to *infer* new knowledge.
- The conceptual map is linked to the topological abstraction of the navigation map and used for resolving linguistic references to entities in the robot's environment (e.g. objects, areas).

T-Box: conceptual knowledge



A-Box: instance knowledge

```
area1:
  is-a 'Area'   (acquired - mapping)
  is-a 'Corridor'  (acquired - place categorization)
  has-a 'obj1'  (asserted - user dialogue)

obj1:
  is-a 'ChargingStation'  (asserted - user dialogue)

area2:
  is-a 'Area'  (acquired - mapping)
  is-a 'Room'  (acquired - place categorization)
  is-a 'LivingRoom'  (inferred)
  has-a 'obj2'  (acquired - vision)

obj2:
  is-a 'Couch'  (acquired - vision)

obj3:
  is-a 'TVSet'  (acquired - vision)
```