Simultaneous Activity Recognition and Monitoring for Robot Assistants

TUM CVPR Group Seminar

Michael Karg

Department of Computer Science
Human Centered Artificial Intelligence Group

March 29th 2012

Outline

1. Robot Assistants in Human Households
2. Spatio-Temporal Plan Representations
3. Simultaneous Plan Recognition and Monitoring (SPRAM)
4. Summary

Outline

1. Robot Assistants in Human Households
2. Spatio-Temporal Plan Representations
3. Simultaneous Plan Recognition and Monitoring (SPRAM)
4. Summary

Robot Assistants in Human Households...

Should:

- Carry out heavy and tedious tasks for humans
- Assist humans in tasks they cannot or do not want to perform
- Carry out tasks self employed

Should not:

- Hinder humans in any way
- Be annoying (vacuum bedroom while human sleeps)

Service Robots in Human Households...

- Need to know what their human partner is doing even without beeing explicitly told
- Need to react adequately to human behavior
- Should learn from observations

Human Belief State Module

The robot should have a module that maintains a belief-state about the activities of its human partner!

Challenges for a Human Belief State Module

- High uncertainties
- World is not static any more
- Human behaviour is hard to model and interpret
- Human might change his mind or perform several tasks simultaneously

Idea: Simultaneous Plan Recognition and Monitoring

Probabilistic framework that keeps track of activities that are likely to be executed and constantly allow for changes.

Cognition-Endabled, Reactive Robot Control

- ROS middleware
- CRAM - Cognitive Robot Abstract Machine for flexible, reliable, and general robot control
- CPL plan language (based on CommonLISP)
- Knowrob knowledge processing system (based on Prolog)

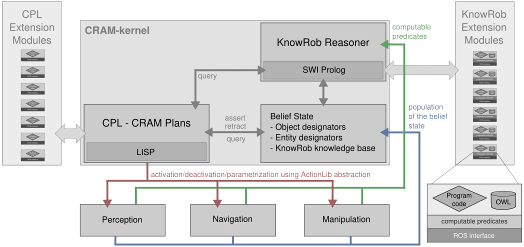


Image courtesy of Michael Beetz / TUM-IAS group

The KnowRob Knowledge Processing System

- Work by Moritz Tenorth et al.
- Tools for knowledge acquisition, representation and reasoning that are tailored to the demands in mobile robotics
- Combines knowledge about the environment, objects, actions etc. obtained from observations or the Web (OpenCyc, WikiHow, ...)
- Knowledge represented Ontolgies using Web Ontology Language (OWL)
- Describe relational knowledge using Description Logics
- Allows queries about e.g. likely storage locations of objects based on the type of object and the container and how to open the specific container

Semantically Annotated Environment Information

- Objects and environment map represented in knowledge base
- Furniture pieces as object instances inherit properties of their type
- Articulation models for opening containers (Jürgens work)
- Spatio-temporal representation of object-poses

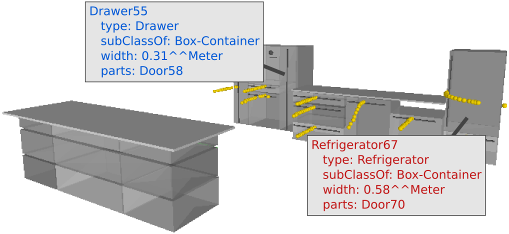


Image courtesy of Moritz Tenorth

Example-query: Where is the Pancake-Mix?

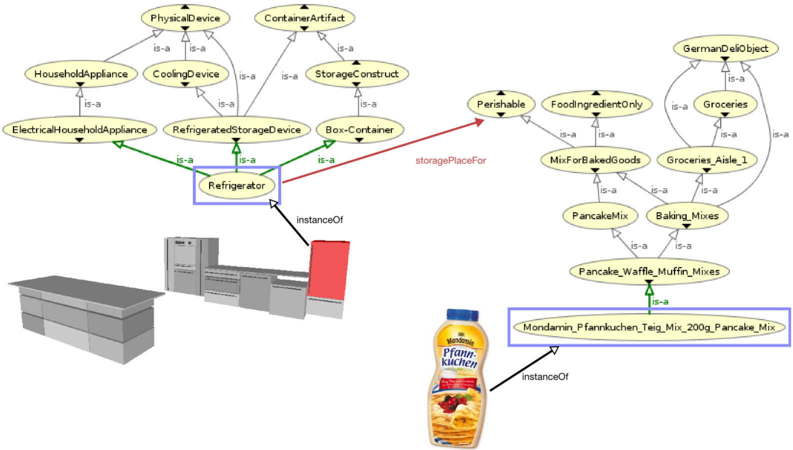


Image courtesy of Moritz Tenorth

Outline

Spatio-Temporal Plan Representations

- Model for human activities based on observation of human task performance
- General, humanlike representation of locations based on semantic environment maps
- Transferable to other environments given a semantic map
- Allow for plan monitoring and -recognition in different environments

Goal:

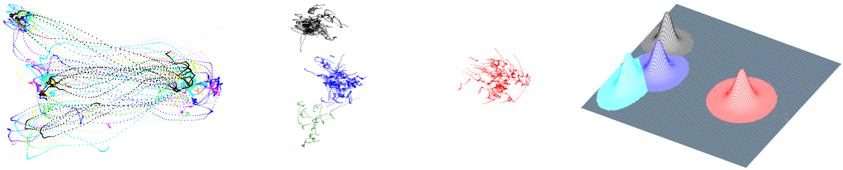
A general, transferable representation of human tasks that allows a robot to explain its observations

The TUM Kitchen Dataset

- Labeled motion-tracking data of humans performing a table-setting-task in a kitchen environment
- 6 objects stored in 3 different locations (cupboard, drawer, stove) plus goal location (table)
- Labels for actions of both hands and body in general

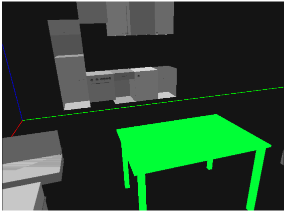
Spatial Model Generation

- Assumption: Human most of the time is standing still while interacting with objects
- Estimate positions where human is standing still and interacts with objects using motion tracking data and labels
- Perform clustering using Expectation Maximization Algorithm



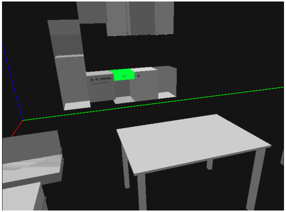
Spatial Model Generation

- Idea: Represent locations relative to furniture objects in the environment
- Assumption: Storage locations of objects known
- Query KnowRob to find storage locations of objects involved in plan
- Put 2D-Gaussians into reference to nearest storage location



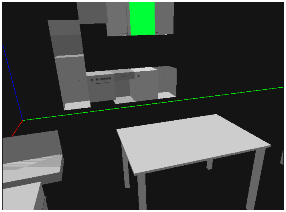
Spatial Model Generation

- Idea: Represent locations relative to furniture objects in the environment
- Assumption: Storage locations of objects known
- Query KnowRob to find storage locations of objects involved in plan
- Put 2D-Gaussians into reference to nearest storage location



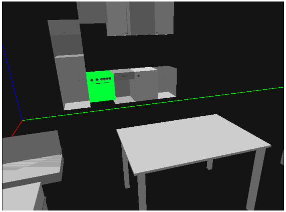
Spatial Model Generation

- Idea: Represent locations relative to furniture objects in the environment
- Assumption: Storage locations of objects known
- Query KnowRob to find storage locations of objects involved in plan
- Put 2D-Gaussians into reference to nearest storage location



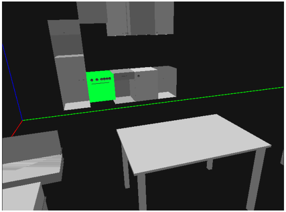
Spatial Model Generation

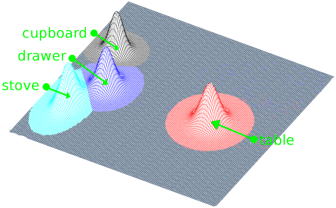
- Idea: Represent locations relative to furniture objects in the environment
- Assumption: Storage locations of objects known
- Query KnowRob to find storage locations of objects involved in plan
- Put 2D-Gaussians into reference to nearest storage location



Spatial Model Generation

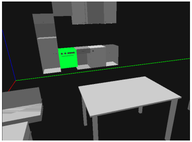
- Idea: Represent locations relative to furniture objects in the environment
- Assumption: Storage locations of objects known
- Query KnowRob to find storage locations of objects involved in plan
- Put 2D-Gaussians into reference to nearest storage location

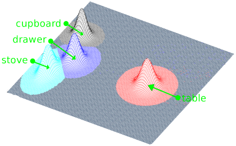




Spatial Model Generation

- Idea: Represent locations relative to furniture objects in the environment
- Assumption: Storage locations of objects known
- Query KnowRob to find storage locations of objects involved in plan
- Put 2D-Gaussians into reference to nearest storage location





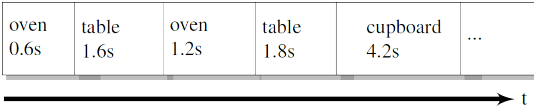
General Spatial Model of locations a human visits during a table-setting task.

Spatio-Temporal Plan Representations (STPRs)

- Representation of human activities based on spatial model using KnowRob-linked locations
- Sequence of n tuples with location l_i and duration t_i:

$$p_n = ((l_1, t_1), (l_2, t_2) \ldots, (l_n, t_n))$$

- Visualization: Timeline-like representation

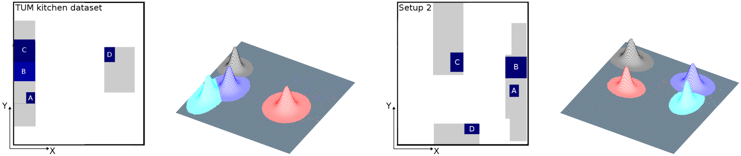


Generation of STPRs

- Analyze human motion tracking data with regards to spatial model
- Create sequences of location/duration tuples

Transferring Spatial Models to Other Environments

- Spatial model can be transfered to other environments given a semantic map and storage locations of objects
- Obtain locations of objects and their orientation from semantic map
- Create gaussians relative to container objects using the learned relations



Outline

1. Robot Assistants in Human Households
2. Spatio-Temporal Plan Representations
3. Simultaneous Plan Recognition and Monitoring (SPRAM)
4. Summary

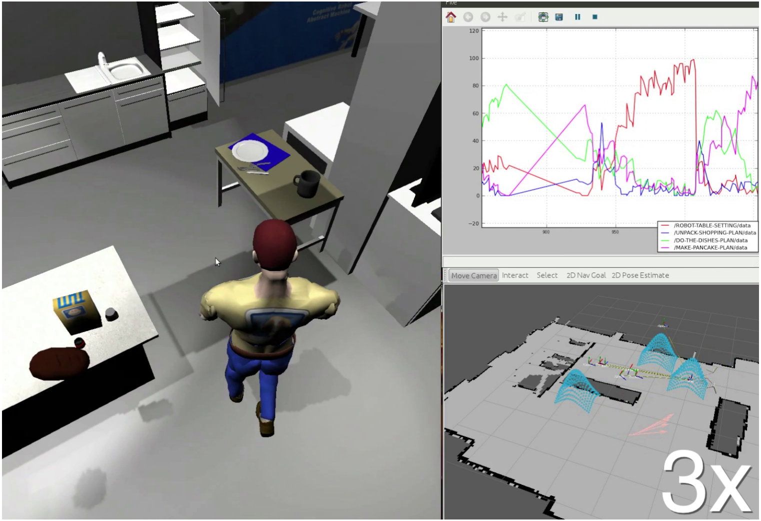
Challenges in Plan Recognition

- High uncertainties
- Only partial observations of objects/human position might be available
- Human might suddenly change its plans or abandom it
- How to detect plan-endings despite only partial observations?

Idea:

Generate a SPRAM module that maintains a posterior probability distribution about human task execution.

Video: Probabilistic SPRAM Module in Action



A High-Level Particle Filter for SPRAM

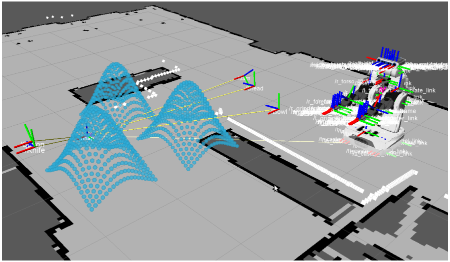
- Model estimation using a High-Level Particle Filter
- Describes posterior probability distribution by a set of particles
- Particles include STPRs (with spatial model) as human task models
- Monte-Carlo based filtering approximates the posterior $p(x_t|z_{1..t})$ by:

$$\int f(x_t)\, p(x_t|z_{0..t})dx_t \approx \frac{1}{P}\sum_{L=1}^{P} f(x_k^{(L)})$$

where x_t are the human activites, $z_i = (location_i, duration_i, objects_i)$ and $f(...)$ represents the weighting function.

A High-Level Particle Filter for SPRAM

- Random Particle Injection prevents degeneration (human might change his plan, plan ending might not be detected)
- Weighting function combines locations, durations, object-detections and overall execution time
- Simultaneous monitoring of most-likely task(s)
- Prediction of places that are likely to be visited by human in the next time



Work In Progress

- Monitoring of several likely tasks
- Combination of STPRs with partial order-models from KnowRob would allow for more elaborate reasoning and improve recognition
- Set up realistic ontology about a “normal” day of a human based on real-world data
- Improve performance using a Relational Particle Filter (Assumption: Obervations conditionally independet which they are NOT!)
- Include paths between places into prediction of places

Outline

1. Robot Assistants in Human Households

2. Spatio-Temporal Plan Representations

3. Simultaneous Plan Recognition and Monitoring (SPRAM)

4. Summary

Summary

- We use Spatio-Temporal Plan Descriptions (STPR) for plan monitoring and recognition in human centered environments
- We set up a module that performs Simultaneous Plan Recognition and Monitoring based on STPRs and semantic environment information
- STPRs can be used accross environments given a semantic map (e.g. from RoboEarth)
- First experiments look promising and there is more to come!

The end

Any questions?

Video: The MORSE Simulator

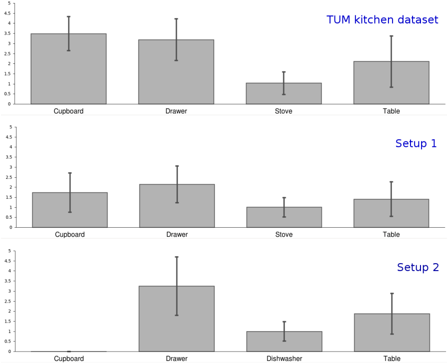
Application Example: Basic Plan Monitoring

- Durations a human spends at places while performing pick and place tasks should be similar in different environements
- Assumption: Durations a human spends at one location depends on amount of manipulation that has to be performed

Question

Are durations a human spends at different types of storage locations comparable?

Application Example: Basic Plan Monitoring



Question

Can we use this information to distinguish a pick and place task from other tasks?

Application Example: Basic Plan Monitoring

- Use durations from TUM Kitchen Dataset as model and calculate confidence value based on durations at storage locations

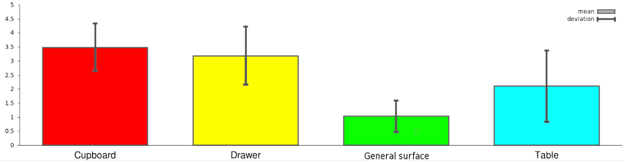


Table setting task:

stove 0.9 s	table 1.2 s	stove 1.0 s	table 1.4 s	cupboard 3.3 s	table 1.9s	...

Cleaning task:

table 1.9 s	stove 3.1 s	drawer 3.3 s	table 2.2 s	...

Application Example: Basic Plan Monitoring

- Use durations from TUM Kitchen Dataset as model and calculate confidence value based on durations at storage locations

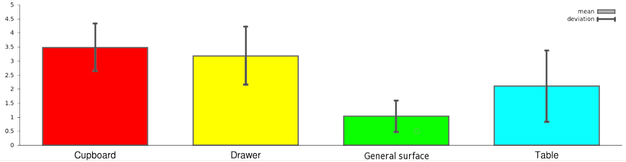


Table setting task:

stove 0.9 s	table 1.2 s	stove 1.0 s	table 1.4 s	cupboard 3.3 s	table 1.9s	...

Confidence: 0.593

Cleaning task:

table 1.9 s	stove 3.1 s	drawer 3.3 s	table 2.2 s	...

Application Example: Basic Plan Monitoring

- Use durations from TUM Kitchen Dataset as model and calculate confidence value based on durations at storage locations

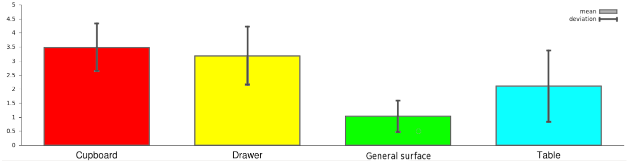


Table setting task:

stove 0.9 s	table 1.2 s	stove 1.0 s	table 1.4 s	cupboard 3.3 s	table 1.9s	...

Confidence: 0.593

Cleaning task:

table 1.9 s	stove 3.1 s	drawer 3.3 s	table 2.2 s	...

Confidence: 0.350

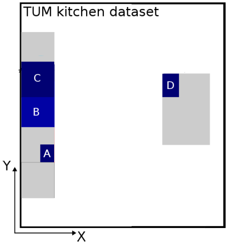
Application Example: Basic Plan Monitoring

- Experiments in two different environments (setup 1, setup 2) using model of TUM Kitchen Dataset
- Recorded motion tracking data of 10 participants performing 3 different tasks:
  - Robot-like table setting
  - Human-like table setting
  - Cleaning task
- Results:

Task	c_p Setup 1	c_p Setup 2
Robot-like table setting	0.524	0.593
Human-like table setting	0.448	0.506
Cleaning task:	0.191	0.350

Application Example: Basic Plan Monitoring

- Use plan patterns to calculate confidence value based in string comparison methods (e.g. Levenshtein distance)



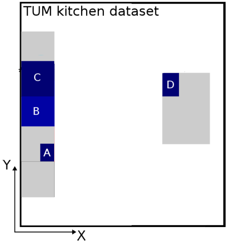
- A: Initial location of placemat and napkin
- B: Initial location of cutlery
- C: Initial location of plate and cup
- D: Goal location

Table setting task:

stove	**table**	**stove**	**table**	**cupboard**	**table**	**...**
0.9 s	**1.2 s**	**1.0 s**	**1.4 s**	**3.3 s**	**1.9s**	

Application Example: Basic Plan Monitoring

- Use plan patterns to calculate confidence value based in string comparison methods (e.g. Levenshtein distance)



- A: Initial location of placemat and napkin
- B: Initial location of cutlery
- C: Initial location of plate and cup
- D: Goal location

Table setting task:

A	**B**	**A**	**B**	**C**	**B**	...

Application Example: Basic Plan Monitoring

Use Generalize Levenshtein Similarity to calculate confidence value:

Table-setting model learned from TUM kitchen dataset:
ADADCDBDBDBDCD

Table-setting task observed in environment 2:
ADADCBDBDBDBDCD

Cleaning-task observed in environment 2:
DACDADBC

Application Example: Basic Plan Monitoring

Use Generalize Levenshtein Similarity to calculate confidence value:

Table-setting model learned from TUM kitchen dataset:
ADADCDBDBDBDCD

Table-setting task observed in environment 2:
ADADCBDBDBDBDCD

Confidence: 0.943

Cleaning-task observed in environment 2:
DACDADBC

Confidence: 0.342

Application Example: Basic Plan Monitoring

Generalized Levenshtein Similarity for 3 different tasks in 2 different environements using model of table setting task in TUM Kitchen environment:

Task	GLS $_{\text{Setup 1}}$	GLS $_{\text{Setup 2}}$
Robot-like table setting	0.982	0.943
Human-like table setting	0.429	0.429
Cleaning task:	0.357	0.340

Conclusion:

We can distinguish different tasks according to their patterns and durations using spatio-temporal plan descriptions!