



MSc - TOPICS IN HUMAN-ROBOT INTERACTION

Research

The research in our group focuses on developing theories, methods, and implementations for human-robot interaction. One of the major challenges we need to address here is how the robot can not only understand what a human user is saying, but also what is being talked *about* – i.e. how to relate dialogue to an understanding of the situation(s) in which the human and the robot find themselves. For more information, including videos, about what we are doing in human-robot interaction in the domains of object manipulation, visual learning, and human-assisted mapping of environments, see [<http://www.dfki.de/cosy/>].

Doing an MSc thesis in human-robot interaction

Robots are fun. When things work, you *see* what's going on. But, robots are also very complicated. Robot perception is noisy, and inaccurate. Software architectures are distributed systems, running on different machines. They include modules covering diverse capabilities such as dialogue processing, planning, vision, and robot control. What this means for doing an MSc thesis in the area of human-robot interaction is simply this: It means team-work, and an openness and willingness to work with people from different scientific backgrounds, possibly located at other places in the world. If you believe you have what it takes, then we offer a stimulating environment in which you can get the support, supervision, and every opportunity to make the most out of your MSc thesis.



Open thesis topics

Below is a list of open thesis topics. Most thesis topics involve a mixture of theory development, linguistic modeling, and implementation on a robot platform (usually in Java). If you are interested, please contact Geert-Jan Kruijff at gj@dfki.de, preferably with the thesis topic(s) you would be interested in.

"Context-sensitive understanding of vague scalar predicates in visually situated dialogue"

When is an object big, or small? when is it high, or low? big, low, slow -- these are all "vague scalars", the interpretation of which is very much dependent on the context in which they are used. the goal of the thesis is to expand on work by DeVault & Stone, to develop an approach for interpreting vague scalar predicates applied to visual objects, used in situated dialogues / human-robot interaction for visual object learning and object manipulation. see also [<http://www.research.rutgers.edu/~ddevault/>].

"Producing deictic gestures to accompany spatial referring expressions in human-robot interaction"

Imagine a robot in a cluttered room. it has just passed inbetween a box, and a wastebin, and is wondering whether it has just gone through a door -- or not. so, at some point, it raises the question: "is there a door here?" [<http://www.dfki.de/cosy/www/media/explorer.y2.html>] without making clear where "here" is supposed to be (the position of the robot? or, this room? or, somewhere nearby the robot?) the human user has no way of knowing what the robot is asking something about, and may thus fail to give a proper answer. the goal here is to develop methods for planning and realizing deictic gestures to accompany spatial referring expressions in human-robot interaction. as deictic gestures the intention is to use a combination of looking (pan-tilt unit), pointing (arm), and turning & positioning (robot pose, movement-to-position). the methods are to be integrated into a dialogue system for human-robot interaction, to be evaluated with users on the comprehensibility of questions and assertions modulo the presence/absence of deictic gestures. possible dialogue domains are visual learning, object manipulation or human-augmented mapping.

"Adaptive interleaving of dialogue- and action planning for manipulation plan execution"

If a robot is given a possibly complex plan for manipulating objects in a scene, the robot should provide feedback to indicate that it has understood what to do. however, such "dialogue grounding" does not need to be done purely verbally -- performing particular initial manipulation actions while producing verbal feedback can also function non-verbal feedback, while at the same time already executing the plan. the goal is (a) to investigate appropriate mixtures of verbal- and non-verbal feedback in manipulation tasks, using a corpus of human-human collaborative dialogue (Baufix corpus), and (b) to develop methods for interleaving dialogue- and action-planning to integrate planned manipulation actions as non-verbal feedback in the production of dialogue grounding moves. because the robot may be interrupted in its action execution (e.g. if it does the wrong thing, or if the situation changes), the approach needs to integrate with a continuous planning model being developed within the project. see also [<http://www.informatik.uni-freiburg.de/~brenner/Publications/brenner:nebel06.pdf>]

"Producing verbal and non-verbal communication engagement acts in human-robot interaction"

How should a robot indicate that it wants to talk to you? or, that it hands over the turn to you? or that it would like to move on, and say goodbye? these are "meta-level" acts that guide how dialogue proceeds as a form of *engagement*. The goal here is to expand on work by Sidner et al to develop methods for planning, and realizing, verbal and non-verbal communication in human-robot interaction to indicate open and closing an engagement, and maintaining or handing over turns. the approach should be evaluated with expert and naive users on the effect of the naturalness (Likert-scale) of dialogue given a use (non-use) of varying forms of non-verbal communication (gaze; posture; deictic gestures) in human-robot interaction in (simple) office assistant scenarios. see also [<http://citeseer.ist.psu.edu/sidner04explorations.html>].

"Comprehending perspectivization in spatial references to locations relative to landmarks"

When referring to an aspect of spatial organization, people do not always adopt an ego-centric perspective. very often, an "allo-centric" perspective is adopted that is centered not on the perceiver, but on something in the environment -- e.g. a salient landmark. the goal is to develop methods for resolving spatial references to locations relative to landmarks, using the robots knowledge about the spatial organization of the environment. the first step is to resolve references to locations relative to landmarks that the robot knows about -- landmarks marked in a map the robot has built up. the second step is to resolve references to locations in known areas, but relative to landmarks that the robot does not know about. this requires the robot to resolve the reference by constructing a (continuous) plan for traveling to the known area, and then performing simple visual search to identify the referred-to location. the final step is to handle references to locations in unknown areas. this requires the robot to construct a continuous plan for traveling to the unknown area (assuming that area can be identified relative to known areas; e.g. "the kitchen [unknown] is at the end of the corridor [known] to your left"), and then perform visual search to identify the referred-to location. the approach is to be evaluated at each stage using a constructed test bed involving simple office assistant tasks ("take the X to the Y and put it R-RELATIVE to the Z"), measuring success in terms of recall (#resolutions construed) and precision (#correct resolutions).

"Stressing what is contextually salient when communicating assertions about spatial organization in a visual scene"

The goal is to expand on recent work by Cassell et al [<http://www.soc.northwestern.edu/justine/>] and Mutlu et al [<http://www.bilgemutlu.com/>] to develop methods for (a) determining a shallow form of information structure for an assertion that the robot needs to communicate, on the basis of a combination of discourse salience and what is salient in the current visuo-spatial scene, and (b) to use that shallow information structure to determine what intonation to use, and where to look, when realizing the assertion. for synthesizing speech with varying intonation, the Mary system is to be used [<http://mary.dfki.de>]. to evaluate the approach, the work is to focus on the production of assertions that describe visual scenes, i.e. collections of objects and their spatial relations, measuring how well expert and naive users are capable of identifying what objects the robot is talking about. of particular interest here is the trade-offs between the use of non-verbal (gaze) and paralinguistic (intonation) means to stress saliency, and the complexity of spatial referring expressions [<http://www.dfki.de/cosy/www/pubs/pdfs/kelleher+kruijff.gre.accoling2006.pdf>]. see also [http://www.bilgemutlu.com/publications/Mutlu_HUMANIDS06.pdf]

"using contrastive intonation in clarification questions for resolving situational ambiguity"

In a clarification question, a robot needs to contrast what it does understand about a situation, with that which it does not. the goal here is to develop methods for planning and realizing contrastive intonation in clarification questions raised in situations when the robot is uncertain about how to interpret something -- for example, "is there a door here?", "is this an box or a ball?", "is the box red or orange?". for synthesizing speech with varying intonation, the Mary system is to be used [<http://mary.dfki.de>]. the resulting methods should be evaluated with expert and naive users on naturalness, and comprehensibility (what is the question about?), possibly using off-line web-experiments.

"Layered short-term scene understanding"

Imagine a robot, capable of picking up objects and taking them from one place to another. when the robot wants to pick up an object on a table, it needs to focus on the table, construct a model (A) of the spatial organization of what is on the table, and pick up the object. next, when moving around the room, it should maintain a model (B) of the current organization in the room -- e.g. the trash bin may have moved from where it was last time. assume the robot needs to drop the object into the trash bin; it needs to go there, focus on the trash bin, basically create yet another model (C) of how the trash bin is positioned relative to the robot, and then drop the object into the bin -- task completed, mission accomplished. the question we can raise here is how models A, B, and C are related. they clearly have a different level of granularity in spatial organization -- a table, a room, the area around a trash bin. yet, they are related -- humans do not forget what they saw on a table, or where the trash bin was when they last saw it. the goal here is to investigate how, in visuo-spatial working memory, the robot can construct and maintain models of visuo-spatial organization at different levels of granularity. for varying granularity the focus is on models that can be anchored on landmarks or identifiable areas. issues like how these models are related among themselves, and how they can be grounded in e.g. a map of the environment, need to be addressed.

"Semantic integration for robust spoken dialogue comprehension"

Automatic speech recognition (ASR) is never 100% perfect: it may wrongly recognize words, "mishearing" what was said, possibly yielding ungrammatical strings as output. as a result, (incremental) parsing of such strings may only be able to yield a collection of partial semantic analyses. the goal here is to develop a method for integrating these partial analyses, using salient structures in the discourse context, categorical knowledge about how meanings can prototypically be combined, and knowledge about salient objects and spatio-temporal organization in the situated context. the resulting approach will be evaluated on spoken-out-loud constructed input (test-bed) and corpus sentences (Home Tour corpus). see also [<http://www.techfak.uni-bielefeld.de/ags/ai/publications/papers/Huwel2006-RSU.pdf>]

"Using situational awareness to improve spoken dialogue recognition"

Automatic speech recognition (ASR) is never 100% perfect: it may wrongly recognize words, "mishearing" what was said, possibly yielding ungrammatical strings as output. the idea here is to expand on work by Purver, and by Roy to use the robot's knowledge of the (visual) situation to resolve competing ASR interpretations to one that is situationally best supported. for example, if the situation contains a ball, but no box, and the top-ranked strings are "this is a ball" and "this is a box", then the idea is to choose for "this is a ball" even when "this is a box" may have a (marginally) higher confidence score. the goal of the thesis is to develop a post-processing module, which takes a word lattice from the ASR engine, and outputs a string which has both a high confidence and which is supported by the situational context. to elaborate the notion of "situational support", the thesis should explore how information from different modalities can be explored -- visual-spatial working memory, but also discourse context, and information about where the human is looking in a scene. see also [<http://www.stanford.edu/~mpurver/papers/purver-et-al06icslp.pdf>] and [http://web.media.mit.edu/~dkroy/papers/pdf/roy_2000_c.pdf].

"Resolving clarification ellipsis in situated dialogue"

Clarification questions are often answered with short phrases in which much of the contextually salient content is simply omitted; e.g. "What color is the box?" "Orange." empirically, the phenomena can be observed in corpora for human-human dialogue (e.g. Brown corpus; Baufix corpus) and human-robot dialogue (Home Tour corpus). the goal is to expand on the work by Purver and Ginzburg to resolve clarification ellipsis in a situated manner, drawing not only on the discourse context but also on what is salient in the situation to resolve ellipsis. the resulting approach is to be implemented and evaluated in situated dialogues in human-robot interaction for visual object learning or human-augmented mapping. see also [<http://www.stanford.edu/~mpurver/papers/pgh01sigdial.pdf>] and [<http://www.dcs.kcl.ac.uk/staff/ginzburg/clar-utt.ps>]

Contact

Dr. ir. Geert-Jan M. Kruijff, gj@dfki.de

For more on our group, see <http://www.dfki.de/cosy/>