FP6-IST-002020

COGNIRON

The Cognitive Robot Companion

Integrated Project

Information Society Technologies Priority

D7.1.1
Specification of the 3 Key-Experiments

Due date of deliverable: 31/12/2004
Actual submission date: 31/01/2005

Start date of project: 01/01/2004 **Duration** : 48 months

Organisation name of lead contractor for this deliverable:
Fraunhofer IPA

Revision: V1
Dissemination level: PU

Executive Summary

This deliverable is the specification document which contains detailed information about the key-experiments (scenarios) of the project. It is the deliverable D7.1.1.: Specification of the three key-experiments, the experimental design of the demonstrations and the partner's role and interactions for the projects duration.
During the activities in RA7 it has been recognized that it is a real challenge to define research outcomes and partner interactions precisely in order to achieve the results desired. Therefore, the 'Cogniron Functions' (CFs) were defined to associate functionalities with key experiments and hence to connect project partners. The first part of this document describes such functions in depth. Hence, it is the largest part and contains already much information on the key-experiments. The second part defines the key-experiments to a selected degree of detail. The last chapter contains the interaction diagram which depicts all collaborations.

Role of the Specification Document in Cogniron

The objectives of the specification document are:

- Specification of the key experiments and function definitions.
- Clarification of collaboration between partners.
- Ensuring a high quality of the implemented research results of RA1 to RA6

Relation to the Key Experiments

The specification document defines the key experiments and keeps tack of the state of the settings.

Contents

1 Specification of the Key-Experiments 4
1.1 Introduction 4
1.2 Description of the Cogniron Functions 5
CF-DLG: Multi modal dialog 5
CF-ROR: Resolving multi-modal object references 8
CF-PTA: Person tracking and detection of attention 10
CF-GR: Gesture recognition 12
CF-IA: Intentionality attribution 13
CF-LOC Dialogue or perception based localization 13
CF-NAV: Navigation based on topological maps 13
CF-SOC: Socially acceptable interaction with regard to space 13
CF-OR: Object recognition and modeling IPA 13
CF-RET: Reasoning about tasks, about its own abilities 13
CF-NHP: Navigation in the human presence 13
CF-MHP: Manipulation in the human presence 13
CF-TBP: Tracking of human body parts for observation 13
CF-ACT: Detection and interpretation of human activities and postures 13
CF-LCT: Learning complex task descriptions 13
CF-RG: Learning to reproduce gestures 13
CF-LIF: Learning important features of a task 13
1.3 Robot Home Tour 13
Description 13
Implementation 13
1.4 The Curious Robot 13
Description 13
Implementation 13
1.5 Learning Skills and Tasks 13
Description 13
Implementation 13
1.6 Interaction Diagram 13
2 Conclusions and Future work 13
3 References 13
3.1 Applicable documents 13

1 Specification of the Key-Experiments

1.1 Introduction

This document is a specification document which contains detailed information about the key-experiments (scenarios) of the project Cogniron (www.Cogniron.org). It is the deliverable D7.1.1.: "Specification of the three key-experiments, the experimental design of the demonstrations and the partner's role and interactions for the projects duration."
The project is organised into six research areas (RAs). To demonstrate the outcomes in an integrated fashion there is another RA, RA7 that is devoted to the set up of so-called key-experiments.
The main goals of the activities in RA7 are:

- Specification of the key-experiments and function definitions.
- Clarification of collaboration between partners.
- Ensuring a high quality of the implemented research results of RA1 to RA6

To reach these goals key-experiments were defined in the first phase of RA6. The following rules have been taken into account:

- There are three key-experiments and three hardware set-ups consisting of a robot, an environment, a script of what is demonstrated and implemented functions which enable the robot following the script.
- Each of the experiments demonstrates a subset of the outcomes of the different working packages of the project. At least two working packages should be shown in one key-experiment.
- There should be at least two partners involved in one set-up.
- Every partner should be involved in at least one set-up.

The three key-experiments are specified in this document: Robot Home Tour, The Curious Robot, and Learning Skills and Tasks.

During the activities in RA7 it has been recognized that the interactions between partners have to be precisely defined in order to achieve the results wanted. Therefore, the 'Cogniron Functions' (CFs) were defined to associate functionalities with key-experiments and hence to connect project partners. The first part of this document describes such functions in depth. Hence, it is the largest part and contains already much information on the key-experiments. The second part defines the key-experiments to a selected degree of detail. The last chapter contains the interaction diagram which depicts all collaborations.

The evaluation issues are only mentioned very coarsely in this document. Since it took more time than expected to define all CFs the evaluation issues will be listed in a second document which is planned for phase two and three in RA7. This will be the evaluation document. Evaluation will be done at two levels: evaluation of each complete key-experiment by user studies and evaluation of each CF by tests on defined data sets.

1.2 Description of the Cogniron Functions

CF-DLG: Multi modal dialog

Description

The multi-modal dialog system processes multi-modal input (e.g. speech and gesture) with the goal of extracting the user's intention. The user intention will be analysed with regard to its role in the current dialog exchange and to the current state of the world. A structured knowledge base needs to be set up for this purpose. Successfully processed results are transformed to system commands and forwarded to responsible modules of the robot system. Non-processable user commands need to be handled systematically.
The dialog module will be responsible for generating multi-modal output - taking social aspects into special consideration - to inform the user on the current state of the processing of the user's command.

State-of-the-Art

Most dialog systems are developed for information services where they act as an interface between the user and a database. Such systems are generally based on finite state machines and slot filling techniques (e.g. [1][2][3]). While this is an elegant and simple solution for information services such approaches do not allow for flexible mixed-initiative dialogs as required by more dynamic human-robot interactions. More sophisticated linguistic and psychological models together with probabilistic models have been used for building more flexible dialog systems (e.g. [4][5]). One of the most influential projects in this field is the TRAINS/TRIPS project ([6][7]) where comprehensive analyses in user intention recognition,
speech understanding, dialog handling and task performance are beingcarried out. However, such models are generally implemented as aninterface to static systems such as databases.
Also, only few systems integrate other modalities than speech. Generally, additional modalities comprise the use of artifical channels such as touchscreen, keyboard or pen-based gestures (e.g. [8][9]). Dialog systems for mobile robots with cognitive abilities and physical functionalities require a more sophisticated user command processing scheme because of the complexity of the information needed for a highly dynamic and situated system such as a robot as opposed to static information as provided by databased for information services. Consequently, communication errors are much less predictable than errors in information systems. Additionally, the robot system itself has to produce complex behaviors that require communication between the dialog system and many other modules. Some navigation robots use dialog systems based on information state change ([10][11]). Dialog systems for service robots interacting with humans in everyday life situations have to be based on a more complex principle due to the complexity of the interaction and the environment. But current implementations are relatively simple and only interpret whole command sentences from users ([12]) or use state-based approaches with relatively simple states ([1][13]).

Literature:
[1] D. Goddeau, H. Meng, J. Polifroni, S. Seneff, S. Busayapongchai. "A Form-Based Dialog Manager for Spoken Language Applications" in: Proc. Int. Conf. Speech and Language Processing , pp 701-705, 1996.
[2] W. Ward, B. Pellom. "The CU Communicator System" in: Proc. IEEE Workshop on Automatic Speech Recognition and Understanding, Keystone Colorado, December, 1999.
[3] D. Spiliotopoulos, I. Androutsopoulos, C. Spyropoulos. "Human-robot interaction based on spoken natural language dialogue" in: Proc. European Workshop on Service and Humanoid Robots, 2001.
[4] J. Cahn, S. Brennan. "A Psychological Model of Grounding and Repair in Dialog" in: Proc. of the Fall 1999 AAAI Symposium on Psychological Models of Communication, 1999.
[5] R. Carlson. "The Dialog Component in the Waxholm System" in: Proc. Twente Workshop on Language Technology (TWLT11) – Dialogue Management in Natural Language Systems, University of Twente, Netherlands, 1996.

[6] J. Allen, B. Miller, E. Ringger, T. Sikorski. "Robust Understanding in a Dialogue System" in: Proc. Assoc. for Computational Linguistics, June 1996.
[7] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, A. Stent. "Towards conversational human-computer interaction", AI Magazine, 22(4), 2001, pp 27-37.
[8] S. McRoy, S. Ali, A. Restifficar, S. Channarukul. "Building intelligent dialog systems", Intelligence 10(1), 1999, pp 14-23.
[9] S. Dusan, J. Flanagan. "A System for Multimodal Dialogue and Language Acquisition", in: Proc. 2nd Romanian Academy Conf. on Speech Technology and Human-Computer Dialogue, Romanian Academy, Bucharest, Romania, 2003.
[10] O. Lemon, A. Bracy, A. Gruenstein, S. Peters. "Information States in a Multi-modal Dialogue System for Human-Robot Conversation", in: Proc. Bi-Dialog - 5th Workshop on Formal Semantics and Pragmatics of Dialogue, pp 57 - 67, 2001.
[11] C. Theobalt, J. Bos, T. Chapman, A. Espinosa-Romero, M. Fraser, G. Hayes, E. Klein, T. Oka, R. Reeve. "Talking to Godot: Dialogue with a Mobile Robot", in Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), pp 1338-1343, Lausanne, Switzerland, 2002.
[12] R. Bischoff, V. Graefe. "Dependable Multimodal Communication and Interaction with Robotic Assistants", in: Proc. IEEE Int. Workshop on Robot and Human Interactive Communicatio (RoMan), Berlin, Germany, pp 300-305, 2002.
[13] J. Fry, H. Asoh, T. Matsui. "Natural dialogue with the JIJO-2 office robot", in: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, vol. 2, pp 1278-1283, 1998.

Advances in Cogniron

The final goal of this function is to enable the robot to understand relatively complex user utterances without imposing inconvenient and unnatural restrictions on the user's conversation styles. This requires the robot to convey its internal state and own intentions in a clear way and to take social aspects such as politeness strategies, social distances etc. into account.
To reach this goal we will combine state-of-the-art strategies for user intention recognition, discourse segmentation, as well as dialog and task handling, that are so far only used in static query systems, with a dynamic robot control system. A second issue will be to extend the unimodal dialog system with other (natural) conceptual modalities (pointing gestures, object recognition etc).
The advances include:

- comprehensive linguistic analysis on single utterances for extraction of user intentions
- handling of dialog exchange based on discourse segmentation technique and common ground building
- integration of cues from multiple conceptual modalities
- multi-modal knowledge representation
- generation of body language
- development tightly coupled with evaluation [KTH]
- close interaction with robot control (i.e. allow for interference with planning) [LAAS]

Test Metrics and Incremental Layers

This CF will be evaluated qualitatively in user studies (by KTH). The results will serve as a base for further design decisions. Milestones for the dialog functionality include:
After 18 months

- basic dialog system with two modalities (speech, gestures)
- interaction with control of basic robot functions (active sensors, movement)
- input analysis of simple and complete command utterances

- simple subdialog strategy (predefined subdialogs)
- speech based system feedback

After 36 months

- integration of more modalities and knowledge sources (e.g. scene model, topological maps, location information, CF-ROR)
- input analysis allows for simple violations of complete command sentence structure (e.g. anaphora, ellipses)
- definition of personality-dependent dialog strategies

By 48 months

- integration of more multi-modal knowledge sources
- complex interaction with robot control (e.g. planning)
- input analysis of incomplete, elliptical utterances
- dynamic discourse segmentation
- multi-modal robot output
- dynamic dialog strategies according to interaction style of user

CF-ROR: Resolving multi-modal object references

Description

For a natural human-robot interaction a mobile robot companion has to support modalities that are used during a human-human interaction. The most efficient modalities to communicate information about the robot's environment are the user's pointing gestures in combination with the user's utterances. This information can be used to gain knowledge about objects that become relevant for the robot during the interaction.
Once the robot companion has learned new object information, it can interact with its environment autonomously. For this learning task a system has to be developed that allows resolving object references and uses a long-term memory for storing object information. Therefore the object resolving system will make use of the modules for pointing gesture recognition (CF-GR) and multi-modal dialog (CF-DLG) as well as object recognition (CF-OR) for detecting known objects. Especially the use of the dialog is necessary to achieve a high acceptance and an easy-to-use interface.
Extracting the visual appearance of unknown object instances will depend on combining pointing gestures with verbal information (e.g., 'the green thing') and an appropriate processing of the visual input (e.g., searching for green image areas). This information will be provided to the object recognition module CF-OR to establish a suitable object representation to allow its recognition using CF-OR during subsequent interactions.

State-of-the-Art

Most approaches in the field of object attention systems are limited to stationary platforms and make use of a rich sensory environment [1,2,3]. These systems focus on the evaluation and the execution of the interaction while neglecting all limitations that are given on a mobile robot (e.g., all sensors onboard the robot, limited computational power, highly dynamical environment). Only few approaches use multi-modal input for resolving object references on a mobile robot. However, these approaches focus either on image processing aspects [4,5] or they have only rudimentary dialog capabilities [6].

Literature
[1] C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, and D. Mulanda. ``Humanoid Robots as Cooperative Partners for People'' Submitted to International Journal of Humanoid Robots. In Review. 2004
[2] C. Breazeal,``Sociable Machines: Expressive Social Exchange Between Humans and Robots''. Sc.D. dissertation, Department of Electrical Engineering and Computer Science, MIT. 2000
[3] A. Utsumi, N. Tesutani, and S. Igi, ``View-based Detection of 3-D Interaction between Hands and Real Objects'', Int. Conf. on Pattern Recognition, 2004
[4] R.E. Kahn, M.J. Swain, ``Understanding People Pointing: The Perseus System'', Int. Symp. on Computer Vision, 1995
[5] R.E. Kahn, ``Perseus: An extensible vision system for human-machine interaction'', dissertation, University of Chicago, 1996
[6] S. S. Ghidary et al., ``Multi-Modal Interaction of Human and Home Robot in the Context of Room Map Generation''. Autonomous Robots. 2002

Advances in Cogniron

The goal of the development of the CF-ROR module is to realize a system for resolving multi-modal object references based on the limited sensor data available on a mobile robot. The approach will use the information from pointing gestures, visual features derived from the images of the robot's camera and verbal input. The system will be capable of dealing with a natural way of interaction, i.e., the different processing components will be closely integrated to allow for a fast and robust system response.
Through controlling the individual modules by the object attention system, modules currently unused can be deactivated to save processing power for the relevant functionalities. Besides resolving pointing to known objects, the learning of unknown object instances by user interaction will be realized. This

will become easier through using a priori knowledge about the environment that was collected by the robot previously (e.g., multiple object views).

Test Metrics and Incremental Layers

The object attention will become more powerful during the development. Besides improvements in the individual components (CF-DLG, CF-GR, CF-OR) the overall attention system will become more flexible with respect to

- distance between pointing hand and referenced object (object outside field of view)
- the richness of verbal descriptions for identifying an object ('green' vs. 'the small green apple').
- the amount of information stored in the long-term memory available for resolving object references

After 18 months

A first implementation of a multi-modal object resolving system using speech and visual input will be developed. The system will be able to resolve references for known object types that are pointed at from a short distance (a few centimeters). The information gained during the multi-modal interaction is stored in a long-term memory for later references and autonomously performed actions.

After 36 months

Dynamical adaptation of the long-term memory will be implemented based on the object information that becomes available without explicit interaction. Furthermore the learning of unknown object instances that are only described by spoken utterances (e.g., the color) will be added. By extending the attention control, references to objects that are currently not in the robot's field of view will be resolved by appropriately aligning the robot's sensors. Furthermore, ambiguities (several objects) should be dealt with, e.g. by asking clarification questions ('Do you mean the red cup?').

By 48 months

A major focus of this stage will be to increase the robustness, especially with respect to changes in environmental lighting conditions and ambiguous spoken utterances. Besides, relations between objects should be incorporated for resolving references ('The red thing left to the apple').

CF-PTA: Person tracking and detection of attention

Description

Person tracking takes image, laser, and sound data and fuses these asynchronous cues in order to locate and track humans surrounding the robot. This multi-modal framework for detection and tracking of persons combines the data from the different sensors based on the percepts extracted from the raw data. These percepts represent the sensor-based counterpart of symbolic information like, e.g., a face. Based on the person tracking, the perceptual information linked to tracked individuals is used to detect persons focussing their attention on the robot. For higher-level human-robot interaction (speech recognition, dialog, ...) this detection of the person's attention is crucial to control which person the robot should attend to.

State-of-the-Art

The visual tracking of humans using surveillance cameras is actively researched nowadays but differs from the sensor setup available on a mobile robot as the onboard cameras usually have a smaller field of view and, most importantly, are not static but moving. Besides the basic tracking of humans, a robot companion also needs to detect the attention of the human. This task cannot be solved reliably by vision alone but also requires to incorporate information from other cues, e.g., acoustic data. Only very few mobile robots have such a multi-modal attention system, for example ROBITA [3]. In a similar setup, we have developed an extensible framework for tracking humans based on a variety of cues. The tracking of humans using multi-modal anchoring based on legs, face, and sound percepts has been shown to already provide good tracking results [1].

Recently, the modular framework has been extended to include percepts representing the torso of a human based on a Gaussian mixture model for the shirt color[2].

Literature

[1] J. Fritsch, M. Kleinehagenbrock, S. Lang, T. Plötz, G. A. Fink, and G. Sagerer. Multi-modal anchoring for human-robot-interaction. In: Robotics and Autonomous Systems, Special issue on Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems, volume 43(2-3), pages 133-147, 2003.

[2] J. Fritsch, M. Kleinehagenbrock, S. Lang, G. A. Fink, and G. Sagerer.
Audiovisual person tracking with a mobile robot. In F. Groen, N. Amato, A. Bonarini, E. Yoshida, and B. Kröse, editors, Proc. Int. Conf. on Intelligent Autonomous Systems, pages 898-906, Amsterdam, March 2004. IOS Press.

[3] Y. Matsusaka, S. Fujie and T. Kobayashi. Modeling of Conversational Strategy for the Robot Participating in the Group Conversation. In Proc. European Conf. on Speech Communication and Technology, pages 2173-2176, Aalborg, Denmark, 2001.

Advances in Cogniron

While the tracking and attention detection is already performing quite robustly, there is still potential for adding more cues in order to make the overall system more robust and cope with more variations in the environment and the human behavior. More importantly, the identity of humans and a more fine-grained representation of humans is still lacking. For detecting and tracking the identity of humans, cues like, e.g., face identification will be incorporated and the tracking framework will be extended to incorporate this information and memorize it appropriately. For example, a human face has a fixed identity while the color of the clothing of a human will be stable only during a single day or shorter time spans.

Besides these additions with respect to the represented information about the tracked humans, the framework will also be extended to include more details of the tracked humans that are provided by the algorithms developed in RA2 (CF-TBP and CF-GR). In this way, the person tracking and attention system will not only have information about which persons are tracked, but also about their current body configuration and, possibly, their current gesturing.

Test Metrics and Incremental Layers

As the functionality of the developed approach will be gradually extended in a qualitative way, there are primarily incremental layers of extended functionality during the development process. Test metrics that apply to the quality of the person tracking and attention system are difficult to specify as this would mainly mean specifying how the humans surrounding the robot behave. Except for the dynamic aspects of humans moving around that the anchoring framework needs to cope with, the approaches used for extracting perceptual data rely on certain assumptions. Some of these assumptions are

- sufficient lighting conditions
- small surrounding noise
- clothing differing from background scenery

While the individual perceptual algorithms will undergo only limited improvement, the fusion of the individual cues in the anchoring framework and the addition of other cues will result in an increased robustness to situations where several of the above listed assumptions fail.

After 18 months

In the first phase this Cogniron function will allow the robot to always keep track of humans in its environment and to focus its attention on a specific person (even by following it) if this person is the communication partner. Used modalities are laser range data (=legs), sound data (=speech) and vision data (=face, torso). During successful tracking, identification data from several cues is accumulated and allows linking the tracked human to known individuals.

After 36 months

Successful tracking and identification is the basis for a more detailed analysis of the human and its body parts performed in CF-TBP. The models of the humans surrounding the robot will be enhanced in order to allow the robot to store more fine-grained information about the current configuration and activities of the humans surrounding it. This information will be provided mainly by CF-TBP and CF-GR. Through incorporating this information, details of the tracked humans can be used for other aspects of the robot companion's functionality. For example, the looking direction of a human is very important for resolving object references in CF-ROR.

By 48 months

It is expected to have at the end of this development a full-featured representation of all humans surrounding the robot including their identity and, if currently observed by sensors, their body configuration and performed activities.

CF-GR: Gesture recognition

Description

Gestures form an important part in human communication and interaction. Humans use gestures to express emotions, to intensify verbal statements and to indicate a position or direction (pointing gestures). Obviously, gestures can be used to command a robot companion and communicate with him in a natural, user-friendly manner. A robot companion should be able to recognize pointing gestures and other communication gestures (like "stop") and react to them in an appropriate way. In the case of pointing gestures it should resolve the direction the user is pointing and possibly the object he is pointing at.

State-of-the-art

First efforts on gesture recognition emerged during the late 1960's. Nowadays there exist many approaches to gesture recognition using Hidden-Markov-Models (HMM's). [1] makes use of stereo-vision to detect pointing gestures and to resolve the direction of pointing. Hand gesture recognition combining the appearance of the hand as well as its motion dynamics is addressed in [2]. An approach taking only into account the dynamic changes of a gesture is shown in [4]. Here the trajectory of colored blobs, recognized as hands, is analyzed by a HMM.
Rebollar et al. [3] use a data glove and a 2-DOF arm-exoskeleton to translate sign language into text. The classification of the signs is done using a hierarchical 3-level decision tree. On and Bowden [6] use a unsupervised clustering method to train a gesture classifier in grey-level images.

Literature

[1] K. Nickel, R. Stiefelhagen: *Detection and Tracking of 3D-Pointing Gestures for Human-Robot-Interaction*. Third IEEE Intl. Conf. on Humanoid Robots – Humanoids 2003, Karlsruhe, Germany, October 1-3, 2003.
[2] G. Ye, J. Corso, G. Hager: *Gesture Recognition Using 3D Appearance and Motion Features.* IEEE Workshop on Real-Time Vision for Human-Computer Interaction, Washington, DC, July 2, 2004.
[3] J. Hernandez-Rebollar, R. Lindeman, N. Kyriakopoulos: *A multi-class pattern recognition system for practical finger spelling translation.* Fourth IEEE Intl. Conf. on Multimodal Interfaces. October 14-16, 2002.
[4] A. Just, O. Bernier, S. Marcel: *Recognition of isolated complex mono- and bi-manual 3D hand gestures.* Sixth IEEE Intl. Conf. on Automatic Face and Gesture Recognition, May 17-19, 2004.
[6] E. Ong, R. Bowden: *A Boosted Classifier Tree for Hand Shape Detection.* Sixth IEEE Intl. Conf. on Automatic Face and Gesture Recognition, May 17-19, 2004.

Advances in Cogniron

The goal of this function is to provide a mapping from user movements and gestures to commands and user intentions. It produces output such as new commands and parameters (e.g. directions) used by the robot companion. To achieve this it relies partly on the tracking of human body parts for observation (CF-TBP).
The advances in this Cogniron function are:

- Definition of a set of gestures to be recognized
- Recognition and classification of pointing and communication gestures
- Resolving the 3D-pointing direction (for pointing gestures)

Test metrics and Incremental Layers

To test the gesture recognition functionality, several pointing gestures will be captured and analysed. These include

- Pointing to objects and locations
- Communication/Command gestures like "stop", "go on", waving etc.

After 18 months
A set of gestures will be selected. The focus will be on gestures easy to recognize and with a clear meaning that is useful to command robot companions. Examples are pointing gestures and command

gestures like *stop*, *go on* etc. For some of the easy gestures there will be methods for recognizing and classifying them. For pointing gestures there will be a coarse estimate for the direction.

After 36 months

The 3D body model obtained by the body part tracking (CF-TBP) will be integrated into the recognition and classification of gestures.

After 48 months

Robust detection of the defined gestures will be achieved and integrated into the key-experiments.

CF-IA: Intentionality attribution

Description

Humans have the natural tendency to attribute intentionality to objects that show self-propelled movements in spaces, convincingly demonstrated e.g. in the famous study by Heider and Simmel in 1944 [1]. The anthropomorphic tendency, that includes the attribution of goals, emotions, and even personality to computers [2] and robots likewise [3] has been exploited recently in a variety of robot designs (e.g. zoomorphic Aibos, or anthropomorphic robots such as Kismet or Leonardo), where appearance and behaviour has been designed in a 'life-like' manner in order to elicit certain interpretations. This CF will investigate, in scenarios relevant to Cogniron and the KE's, people's attribution of intentionality to robots, and in return, will study how the robot's behaviour can be shaped to express intentionality in terms of appearance and spatial, non-verbal behaviour.

State-of-the-Art

Recently, psychological studies have systematically studied people's perceptions of robots, e.g. [4,5,6,7,8]. However, a detailed model of how appearance and behaviour of robots shape people's attribution of intentionality and other internal states to a robot is still an open and challenging research issue. Psychological approaches in this area are very promising, but also time consuming. Within Cogniron, the situation is complicated by the fact that we don't have one common experimental platform: the platforms used within the consortium basically cover the whole range from machine-like to human-like platforms. Thus, any contributions of this CF will have to be specific wrt particular robot platforms, and particular behaviours. A comprehensive investigation towards a generic model of intention attribution is out of the scope of Cogniron. However, we will aim at a more focussed approach, investigating a small subset of available robotic platforms, and investigating a limited set of behaviours relevant to the KE's. Carefully designed user studies will identify what attributions certain robot appearances and behaviours elicit. In a later stage relevant aspects of the robot's appearance and behaviour that can be employed by the robot in order to elicit intention attribution will be identified.

Literature

[1] Heider, F. and Simmel, M. (1944) An experimental study of apparent behaviour. *American Journal of Psychology* 57: 243-259.

[2] Reeves, B., Nass, C., (1996) The media equation: How people treat computers, television, and new media like real people and places. CSLI Publications.

[3] Dautenhahn, K (2002) Design Spaces and Niche Spaces of Believable Social Robots. Proc. 2002 IEEE Int. Workshop on Robot and Human Interactive Communication ROMAN 2002, 25-27 September, Berlin, Germany, pp. 192-197, IEEE Press.

[4] Friedman, B., Kahn, P. H. (Jr.), Hagman, J. (2003) Hardware companions? What online AIBO discussion forums reveal about the human-robotic relationship, Digital Sociability, vol. 5, pp. 273-280, 2003.

[5] Khan, Z (1998) Attitude towards intelligent service robots, NADA KTH, Stockholm 1998.

[6] Scopelliti, M., Giuliani, MV, D'Amico, AM & Fornara, F. (2004) If I had a robot at home.... Peoples' representation of domestic robots. In Designing a more inclusive world, S. Keates, J. Clarkson, P. Langdon, and P. Robinson, Eds. Cambridge, UK: Springer, pp. 257-266.

[7] Kahn, P.H. (Jr.), Friedman, B., Perez-Granados, D. R., Freier N. G. (2004) Robotic pets in the lives of preschool children, Proc. 2004 Conference on Human factors and computing systems, 1449 - 1452

[8] Woods, S., Dautenhahn, K., Schulz, J. (2004) The Design Space of Robots: Investigating Children's Views , Proc. IEEE Ro-man 2004, 13th IEEE International Workshop on Robot and Human Interactive Communication September 20-22, 2004 Kurashiki, Okayama Japan

Advances in Cogniron

The goal of this CF is to advance the robot's ability to express intentional behaviour, important e.g. in situations where the robot needs to initiate or maintain interaction or attract the user's attention. This ability will enhance the robot's range of socially acceptable behaviour. Note, expression of intentionality can include a variety of verbal, non-verbal as well as spatial interaction. This particular

CF will focus on spatial, non-verbal means of robot behaviour and how these can be used to express intentionality in interaction with humans.
Specific advances in this Cogniron function are:

- Integration of results from different user studies covering a larger spectrum of robot appearance and behaviour. A comparative approach will allow to go beyond existing studies that usually focus on one particular robot and a limited set of behaviours.

- Identifying correspondences between the robot's behaviour, the particular context and task environment of the study, the robot's appearance etc. on the one hand, and how users attribute intentionality to robots on the other hand

- Developing guidelines for the design of robot behaviours so that the robot can express intentionality and thus become more "readable" to users

- Development of a conceptual model that comprises psychologically relevant parameters of robot behaviour and appearance, and user characteristics with regard to intention attribution.

Test Metrics and Incremental Layers

The robot's expression of intentionality will be studied in user studies that will also inform this CF. Studies will be evaluated with methods including questionnaires and observational analysis. Comparisons of results of the different studies will inform this CF which is linked to CF-SOC.
After 18 months
Identification of relevant parameters of robots' behaviour and appearance that have shown to be important in user studies carried out during 2004, and that can inform work in the KE's, based on literature research and user studies performed within Cogniron during 2004.
After 36 months
Development of design guidelines for robots that can express intentionality efficiently (and socially acceptable). This will result in a conceptual model of expression of intentionality in robots relevant for selected scenarios used in the KE's (note, due to the scope of this research area, the model will inevitably be limited to certain robot platforms, certain sets of behaviours, certain contexts etc.)
By 48 months
Refinement of the model.

CF-LOC Dialogue or perception based localization

Description

This function will take care of the estimation of the robot position using its perception system, possibly aided by human interaction. Localization is based on an internal representation of its (indoor) environment. The representation must be learned by means of an exploratory learning process where human feedback plays a role. This addresses fundamental questions of scene understanding, which include object recognition, extraction of relationships between objects including their temporal properties.

State of the art

Traditionally "metric" models have been used to represent the environment. Such models represent geometric properties of the environment, such as occupancy grids or polygonal representations of free space [4][3][15][10][1]. These models fit very well with the conventional sensors in the robot, which mostly measure the range to objects. A detailed overview on such methods is given in [2]. Metric maps usually scale badly with large environments and are not globally distinct if based on occupancy grids or simple geometric features. Topological maps [7] usually scale well, but are not well suited for precise local tasks. In this research activity, we want to develop and implement a hierarchical representation of space, combining local metric maps with global topological representation [9]. Currently the use of vision systems on mobile robots is commonplace. A natural extension to the `metric' representations is to use a full CAD model [13]. In contrast to such explicit modelling of the environment it is possible to make an implicit model, for example by representing the relation between the sensor data and robot pose, for example by a function approximator or by storing a set of training samples in a database. For vision sensors such an approach is called 'appearance modelling', and has shown to work well in many cases [8]. Based on such a model a probabilistic framework can be used to localize the robot. Another way to represent the environment is by finding a set of localized landmarks [11][12][14][6]. Such landmarks may be derived from range data or may be derived from vision. Also for such a model probabilistic methods are studied to localize the robot [5][1]. The problem of finding the training set, or finding the locations of the landmarks has got an enormous amount of attention the last years. So-called `SLAM' (Simultaneous Localization and Mapping) approaches are able to build a representation (appearance or landmark based) while the robot operates.

Literature

[1] Arras, K.O., Castellanos, J.A., Schilt, M. and Siegwart, R. Feature-based multi-hypothesis localization and tracking using geometric constraints. Robotics and Autonomous Systems, 1056 (2003) 1-13.

[2] J. Borenstein, H. R. Everett, and L. Feng, Navigating Mobile Robots: Sensors and Techniques, A. K. Peters, Ltd., Wellesley, MA, 1996.

[3] Castellanos, J. A., M. Devy, et al. (2000). Simultaneous Localization and Map Building for Mobile Robots: A Landmark-based Approach. IEEE International Conference on Robotics and Automation, (ICRA), SanFrancisco, 2000

[4] R.Chatila, J-P. Laumond. Position referencing and consistent world modelling for mobile robots. IEEE International Conference on Robotics and Automation (ICRA), Saint-Louis (USA), March 1985.

[5] Jensfelt,P., Wijk,O., Austin, D.,Anderson, M. Experiments on augmenting condensation for mobile robot localization, IEEE International Conference on Robotics and Automation, San Francisco, CA, April 2000, pp. 2518-2524

[6] Hayet J-B, C. Esteves, M. Devy, F. Lerasle. Qualitalitative modelling of indoor environments from visual landmarks and range data. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'2002), Lausanne, Swuitzerland, Oct. 2002, pp.631-636.

[7] Kuipers, B. J. and Y. T. Byun. A qualitative approach to robot exploration and map-learning. Workshop on Spatial Reasoning and Multi-Sensor Fusion, Los Altos, CA, USA, Morgan Kaufmann (1987).

[8] B.J.A. Kröse, N. Vlassis, R. Bunschoten, and Y. Motomura. "A probabilistic model for appearance-based robot localization", Image and Vision Computing, 19(6), pp 381-391, April 2001.

[9] Lamon, P., Nourbakhsh, I., Jensen, B. and Siegwart, R. (2001) Deriving and Matching Image Fingerprint Sequences for Mobile Robot Localization. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2001), Seoul, Korea, May 21-26

[10] Leonard, J.J., H.F. Durrant-Whyte, et al. (1992). Dynamic Map Building for an Autonomous Mobile Robot, The International Journal of Robotics Research 11(4).
[11] M.J. Mataric. Integration of representation into goal-driven based-based robots. IEEE Trans. on Robotics and Automation, 8(3):304--312, 1992.
[12] Pierce, A and B. Kuipers: ``Learning to explore and build maps". Proc. AAAI-94, AAAI/MIT Press 1994.
[13] M. Schmitt, M. Rous, A. Matsikis, and K. Kraiss. Vision-based self-localization of a mobile robot using a virtual environment. Proc. 1999 IEEE Int. Conf. on Robotics and Automation , pages 2911--2916, Detroit, USA, 1999.
[14] Stephen Se, David Lowe, and James J. Little. Vision-based mobile robot localization and mapping using scale-invariant features. submitted to ICRA-00, 2000.
[15] Thrun S., Burgard W., Fox D., "A Probabilistic Approach to Concurrent Mapping and Localization for Mobile Robots", Autonomous Robots, KluwerAcademic Publishers, Vol. 31/5, 1998, 1-25

Advances in Cogniron

In the Cogniron project we will build hierarchical representations of space which are used for robot localization. We will develop methods for categorization of space, both sensory driven and driven by interactions with humans. The consistency between representations at different levels has to be kept. The representations will be used in a probabilistic localization.

Test metrics and Incremental Layers

The localization will gradually made more difficult. We will start with static scenes in which illumination conditions remain constant. Then gradually we will move to larger environments with dynamic objects and changing light conditions

After 18 months

The navigation module will use both metric and appearance information to localize. Different methods for categorization will have studied and compared. A database of sensory patterns for each location in space will be available for the users. Robustness for dynamic environments and illumination differences will be studied.

After 36 months

The second period we want to extend the representations with understanding of object, and study whether a categorization and localization can be achieved which is more robust to dynamics and changes in the environment

After 48 months

Test in realistic situations have been carried out

CF-NAV: Navigation based on topological maps

Description

This function will take care of the navigation of the robot, given there is some form of internal representation, and given the fact that there are sensors for monitoring.

State of the art

Traditionally "metric" models have been used to represent the environment and many models for goal-directed navigation and obstacle avoidance have been presented, see for example [1]. In the project we will mainly focus on the *representations* of space, and not so much on the navigation. This means that we will use mainly standard planning and obstacle avoidance techniques and adapt them to the representations which are currently developed.

Literature

[1] J. Borenstein, H. R. Everett, and L. Feng, Navigating Mobile Robots: Sensors and Techniques, A. K. Peters, Ltd., Wellesley, MA, 1996.

Advances in Cogniron

In the project we will mainly focus on the representations of the environment and investigate navigation methods only related to changes in representations. For example, we will study how hierarchical representations will affect the planning strategy.

Test metrics and incremental layers

The navigation will gradually made more difficult. We will start with static scenes in which illumination conditions remain constant. Then gradually we will move to larger environments with dynamic objects and changing light conditions

After 18 months

A navigation module will be described which uses the hierarchical representation developed in WP5.1. For obstacle avoidance, of-the-shelve methods will be use.

After 36 months

If a new environment representation is used, the navigation module had to be adapted and tested

After 48 months

Test in realistic situations have been carried out.

CF-SOC: Socially acceptable interaction with regard to space

Description

A robot companion will operate in a human-inhabited environment, e.g. the home. What is more, a lot of its functionalities will depend on close contact/interaction with humans. Thus, in order to be acceptable to human users, it is vital to investigate requirements and constraints on how the robot should interact with and behave towards humans. Here, people's individual preferences that might dynamically change over time, social norms and conventions, the robot's role in the home, and many other factors are likely to shape its acceptability. Studying and designing socially acceptable behaviour for autonomous robot in a home scenario is a huge challenge, and cannot be addressed in total within Cogniron. Thus, with regard to the KE's, we identified an area of research that is particularly relevant to KE1 and KE2, and that can realistically be addressed within the project of the project: social spaces. Studies will investigate distances that are comfortable to human subjects, depending on the particular context, task, and other environmental factors. Results from different user studies, performed in different laboratories with different robot platforms and focussing on different KE's, will be compared carefully in order to extract common or invariant aspects that allow to derive a set of "social interaction rules". Given the robot's perception of e.g. the location of a human, information about his/her individual preferences, personality, and other context information, these heuristics will provide the robot with default settings for controlling its spatial relationships to humans. In a later stage, these default settings will be subject to adaptation and learning mechanisms so that the robot is able to flexibly cope with dynamic aspects in the (social) environment. Note, "default settings" need to be understood in terms of the above mentioned "social interaction rules", or "maxims" (similar to Gricean maxims known in linguistic communication), suggesting how the robot should move, position and orient itself when approaching human users. Thus, they are not fixed and prespecified settings, but recommendations for robot spatial behaviour in human-inhabited environments relevant to the KE's.

State-of-the-Art

Investigating social spaces for human-sized, mobile robots in a home scenario is a quite recent area of research. Social spaces in human-human interaction have been studied since E.T. Halls' pioneering work in the 1960ies in social sciences [1,2]. Research in virtual environments has studied the role of approach/orientation towards/and distances of virtual agents (cf. [3]). However, only over the past few years the issue of social spaces has been emphasized in robotics. Robotic platforms have become available that allow applications in everyday environments, which has lead to a number of systematic long-term user studies in order to investigate people's reactions towards such robots [4,5,6]. Similar to human-human interaction, robots that interact with people need to be able to control their distances in order to maintain a comfortable zone that facilitates communication and interaction. Note, these distances should not be considered in a purely static sense. The dynamics of moment-by-moment spatial adaptation need to be studied carefully, where both robot and human are moving and able to change their posture, gaze direction and speed as they approach a certain position relevant to the task. This also includes cases when users try to communicate with the robot to make it adapt spatially. Work in this area is related to other work in the project addressing verbal/linguistic communication. Common underlying principles of recipient design, where interactants plan and shape their actions according to the actions of the interaction partner, are a common theme in the pragmatics of both linguistic, non-verbal, as well as spatial interaction between human and robot. Different from other activities in the project (cf. RA1), this CF will focus primarily on spatial interaction, such as keeping distances, speed and character of movements, and bodily orientation towards others who are present. These are important aspects of social interaction that do not necessarily form part of a strictly communicative act, but are crucial in regulating interactions, as well as in preparing for communicative acts.

Literature

[1] Hall, E.T. (1966) The Hidden Dimension: Man's Use of Space in Public and Private. The Bodley Head Ltd, London, UK.

[2] Hall, E.T. (1968) Proxemics. Current Anthropology 9(2-3): 83-108.
[3] Benford, S. and Fahlén, L. (1993) A spatial model of interaction in large virtual environments. Proceedings of the Third European Conference on Computer Supported Cooperative Work (ECSCW'93), Milano, Italy, September 1993.
[4] Severinson-Eklundh, K., Green, A., Hüttenrauch, H. (2003) Social and collaboration aspects of interaction with a service robot. Robotics and Autonomous Systems 42:223-234.
[5] Kanda, T., Hirano, T., Eaton, D., Ishiguro, H. (2004) Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial," *Journal of Human Computer Interaction (Special issues on human-robot interaction)*, 19(1-2): 61-84.
[6] Robins, B., Dautenhahn, K., te Boekhorst, R., Billard, A. (2004) Effects of repeated exposure to a humanoid robot on children with autism. In S. Keates, J. Clarkson, P. Langdon and P. Robinson (Eds.) Designing a More Inclusive World, Springer Verlag, London, pp. 225-236.

Advances in Cogniron

The main aim of this function is to allow the robot to control spatial distances towards human interactants in an acceptable manner, i.e. in a way that is comfortable to the user. This function will enhance the *social capabilities* of the robot, which will complement its technical capabilities. State-of-the-art user studies will inform the design of this CF. It ca not be expected that our CF will generalise across all possible context, robotic platforms, and scenarios for a cognitive robot companion, thus the emphasis on particular settings used in KE's 1 &2.
The expected advances that this function can add to the robot's cognitive abilities:

- Socially acceptable control of distances and approach in a following scenario, involving human-robot dialogue.

- Socially acceptable control of distances and approach in scenarios where a) human and robot moving in a confined space, and b) the robot assisting a human.

- Socially acceptable adaptation to a user's individual preferences

- Socially acceptable techniques of engaging in interaction by means of identifying a human's willingness to interact

- Learning and adaptation mechanisms that allow the robot to adjust its social behaviour

Test Metrics and Incremental Layers

This CF will be tested in user studies, that in return also inform the design of this function. User studies will evaluate the (inevitably subjective) acceptability of the robot's behaviour with standard methods such as questionnaires, behavioural analysis etc, where appropriate.
Incremental Layers:
After 18 months
Initial default settings ("social rules" or "maxims") for socially acceptable behaviour with respect to social distances as relevant to KE1 and KE2 related scenarios, derived from literature studies on human-human interaction and user studies performed within Cogniron in 2004.
After 36 months
Refined settings of the above.

By 48 months
Adaptation of the robot's social behaviour with respect to social spaces. The final version of this CF will allow the robot's social behaviour to adapt to different contexts and/or users. Note, personalization and adaptation, based on long-term human-robot relationships and

repeated exposures will occur on a time scale that might not be suitable for brief demos. However, experimental data and videos will provide sufficient scientific evidence and support exemplifying this CF.

CF-OR: Object recognition and modeling IPA

Description

The cognitive robot companion should be able to identify and localize and manipulate objects which have relevance for task execution. A recognition function should be robust against various sources of noise and fast in calculation. Another important aspect of the recognition system is that the robot must be able to generate object models autonomously since it is usually not known which objects the robot will face in advance. The robot can use its manipulation abilities to produce different percepts of the object from different viewpoints. The Cogniron function will be developed on the basis of two different sensors: colour camera and depth camera.

The Cogniron function CF-OR is also meant to be a more general function which involves autonomous categorization and association of the objects with functions. However, the following descriptions are related to the teach-in process of objects because this is the basis for the other functions.

Learning of object representations through linguistic interaction. In order to acquire object representations that agree with human concepts, we need to extract concepts from the user. A natural way to do this is through linguistic interactions, where the robots listens to human naming objects and the robot inquiring the human about their correct linguistic categories for objects.

State-of-the-Art

The are numerous methods to analyse image content. Recently, there have been explorations using pure learning approaches to appearance-based object recognition [1]. Also, there are approaches which represent objects based on key-point features which are invariant against perspective transformations [2] learning algorithms, such as support vector machines (SVM) [3] or the hierarchical discriminant regression algorithm (HDR) [4] work in high dimensional feature space and are applied directly to images. Work in the field of active perception or active vision [5] stresses the aspect of improving perception through action. One example is active segmentation [6] which is used to segment unknown objects against the background using the manipulation abilities of the robot.

Literature

[1] H. Murase and S. K. Nayar, “Visual learning and recognition of 3-D objects from appearance” *International Journal of Computer Vision*, vol. 14, pp. 5-24, 1995

[2] D. G. Lowe, “Distinctive image features from scale-invariant keypoints” *International Journal of Computer Vision,* 60, 2 (2004), pp. 91-110.

[3] M. Pontil and A. Verri, “Support vector machines for 3-D object recognition” *IEEE Trans. Pattern Analysis and Machine Intelligence,* 20(6), 637-646, 1998.

[4] Weng, J., “Incremental Hierarchical Discriminant Regression for Online Image Classification” *Sixth International Conference on Document Analysis and Recognition* (ICDAR '01), September 10 - 13, 2001.

[5] J. Aloimonos and A. Badyopadhyay, “Active Vision” *IEEE First Int. Conf. on Computer Vision*, pp 35-54, June 1987.

[6] Paul Fitzpatrick. Object Lesson: discovering and learning to recognize objects. *Proceedings of the 3rd International IEEE/RAS Conference on Humanoid Robots, Karlsruhe*, Germany, October, 2003.

Advances in Cogniron

The overall goal of this functions is to make the robot be able to interactively construct models for objects such that it can manipulate them or use them. To do this, latest advances in appearance-based and feature-based object recognition as well as learning theory and active methods will be merged into a single framework.

The advances in this Cogniron function build on the aspects described above. They are:

- Integration of feature approaches and modern learning algorithms into a unified framework.

- Combination of colour and depth sensors: The depth sensor estimates the distance of the object. The search strategy, which incorporates the learning algorithm, can be tuned (varying the size of the search-window) to that distance.

- The third increment to the current state-of-the-art is the use of active strategies for object model and property acquisition. That means, that the robot can get "familiar" with objects though manipulation.

- The last goal of this Cogniron function is also to integrate other sensory-motor data than the vision and depth information to the objects model. This data can be tactile information that result from interaction with the object. This may be extended to the sensory-motor-based development of simple object physics.

Test Metrics and Incremental Layers

To test the recognition system there will be a defined set of different objects with varying properties.

- Colour and other surface properties (e.g. refulgence)

- Shape, size and shape influencing properties (e.g. kinematics)

Also environmental parameters will be varied:

- Lightening conditions

- Different set-ups of the objects incorporating occlusions and other sources of uncertainty

Incremental Layers:

After 18 months

Multimodal (color & depth) object detection and pose estimation of objects related to the KEs with a simple procedure to generate object models by physical interaction with the objects

After 36 months

Advanced techniques for active acquisition of object information related to their appearance in shape & colour from all perspectives. Self-supervised categorization of objects.

By 48 months

Acquisition of object models taking into account tactile and other information available through manipulation of the object. Learning objects attributes. Self-supervised categorization of objects and association with object related functions.

CF-RET: Reasoning about tasks, about its own abilities

Description

This function aims at integrating the other robot functions developed in the project and at providing a framework that will allow the robot to reason about its functions and its abilities to perform its tasks (including human-robot interactive tasks) in a given environment.
Explicit handling of uncertainty is essential for the robots we envisage to build. We aim to develop a representation of uncertainty and an inference model that will serve as a basis to build, in a coherent fashion, the various representations that will be used by the robot: the environment model (and more generally the context as it will be perceived by the robot including human activity), the skills and tasks (pre-programmed or learned).

State of the art

There are numerous contribution in the literature linked to architectural aspects, high-level control and more generally robot decisional issues. Several systems and approaches based on programming by including learned task descriptions as atomic actions into the planning algorithm.

Literature

[1] R. Alami, R. Chatila, S. Fleury, M. Ghallab, F. Ingrand, An architecture for autonomy. International Journal of Robotic Research, Vol.17, Nr 4, pp.315-337, April 1998
[2] R. Volpe, I. Nesnas, T. Estlin, D. Mutz, R. Petras, and H. Das. Claraty: Coupled layer architecture for robotic autonomy. Technical Report, Jet Propulsion Laboratory, Dec. 2000.
[3] F.Ingrand, R.Chatila, R.Alami. An architecture for dependable autonomous robots 1st IARP/IEEE-RAS Joint Workshop on Technical Challenge for Dependable Robots in Human Environments, Seoul (Korea), May 2001
[4] W.Paquier, R. Chatila. An architecture for robot learning. 7th International Conference on Intelligent Autonomous Systems (IAS-7), Marina del Rey (USA), 25-27 Mars 2002, pp.252-255
[5] S. Das, B. Grosz and A. Pfeffer. Learning and Decision-Making for Reconciliation. In the Proceedings of the First Joint Conference on Autonomous Agents and Multi-agent Systems, Bologna, Italy. July 2002.
[6] D. Sullivan, B. Grosz, and S. Kraus. Intention Reconciliation by Collaborative Agents. In Proceedings of the Fourth International Conference on Multi-Agent Systems (Boston, Massachusetts), IEEE Computer Society Press, pp. 293-300.
[7] M. E. Pollack, Planning Technology for Intelligent Cognitive Orthotics, 6th International Conference on AI Planning and Scheduling, April 2002.
[8] R. Alami. A Multi-Robot Cooperation Scheme Based on Incremental Plan-Merging In, R. Hirzinger and G. Giralt (Eds), Robotics Research : The Seventh International Symposium, Springer Verlag, 1996.
[9] S. Bothelo, R. Alami, Robots that cooperatively enhance their plans DARS 2000, October 4-6, 2000, Knoxville, Tennessee, USA
[10] J. Pineau, G. Gordon, and S. Thrun. Policy-contingent abstraction for robust robot control. In Proceedings of the Conference on Uncertainty in AI (UAI), Acapulco, Mexico, 2003.
[11] J.L. Fernández, R. Simmons, R. Sanz and A.R. Diéguez A Robust Stochastic Supervision Architecture for an Indoor Mobile Robot,. In Proceedings of the International Conference on Field and Service Robotics (FSR 2001) Espoo Finland, June 2001.

Advances in Cogniron

The control architecture principles will be based on the state of the art. Our efforts will be devoted to two aspects that are of particular importance in our case: the learning and communication abilities, and their consequence on the overall architecture (inference model, representation and programming). Indeed, it will be necessary to export the essential supervision data (e.g. current goals, tasks, current task/subtask hierarchy, main control parameters...) that may be used in cooperative problem solving or to exhibit robot intentions.

Besides, we aim to build a control architecture and inference mechanisms that are able to scale to realistic problems and situations.

Test Metrics and Incremental Layers

After 18 months
Preliminary design of an architecture for cognitive robot
After 36 months
An implemented architecture with a decisional kernel.
After 48 months
Demonstration in integrated experiments

CF-NHP: Navigation in the human presence

Description

Objectives

The objective is to study how a mobile robot can plan and execute motions in close proximity to a human, so that the human considers the robot's behaviour to be safe and comfortable.

Description

We will develop new methods and algorithms that will allow the robot to compute and perform safe and socially acceptable motion for a mobile robot in an environment populated by humans.

The work will be based on motion planning and execution techniques for robots and humanoids as well as sensor-based motion planning techniques. Besides, we would like to take into account, when selecting relative human-robot configurations and motions, criteria that favour feeling of comfort and non intrusion of the robot. This will be based on the user studies perfomed by partners in Cogniron.

State of the art

While several authors propose motion planning or reactive schemes, there is no contribution that tackles globally and systematically the problem as we propose to do.

Literature

[1] Dautenhahn, K., Ogden, B., Quick T. (2002) From Embodied to Socially Embedded Agents -Implications for Interaction-Aware Robots, Cognitive Systems Research 3(3), pp. 397-428.

[2] Matsusaka, Y., Fujie, S., Kobayashi, T. (2001) Modelling of Conversational Strategy for the Robot Participating in the Group Conversation. Proc. ISCA-EUROSPEECH2001, pp. 2173-2176.

[3] Bobick, A. F. (1996) Computers seeing action. Proc. British Machine Vision Conference, Edinburgh, Scotland, September 1996.

[4] Benford, S. and Fahlén, L. (1993) A spatial model of interaction in large virtual environments. Proceedings of the Third European Conference on Computer Supported Cooperative Work (ECSCW'93), Milano, Italy, September 1993.

[5] Hall, E.T. (1966) The Hidden Dimension: Man's Use of Space in Public and Private. The Bodley Head Ltd, London, UK.

[6] Hall, E.T. (1968) Proxemics. Current Anthropology 9(2-3): 83-108.

[7] Nakauchi, Y., Simmons, R. (2000) A Social Robot that Stands in Line. In Proc. of the Conference on Intelligent Robots and Systems (IROS), Takamatsu Japan, October 2000.

[8] Fong, T., Nourbakhsh, I., Dautenhahn, K., (2003) A Survey of Socially Interactive Robots, Robotics and Autonomous Systems 42(3-4), 143-166.

[9] R. Alami. A Multi-Robot Cooperation Scheme Based on Incremental Plan-Merging In, R. Hirzinger and G. Giralt (Eds), Robotics Research : The Seventh International Symposium, Springer Verlag, 1996.

[10] D. Hähnel, D. Schulz, W. Burgard. Map Building with Mobile Robots in Populated Environments. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2002.

[11] R. Alami, T. Simeon, K. Madhava Krishna. On the influence of sensor capacities and environment dynamics onto collision-free motion plans. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'2002), Lausanne (Suisse), Oct. 2002.

[12] R. Chatila, R. Alami, T. Simeon, J. Pettre, L. Jaillet Safe, reliable and friendly interaction between humans and humanoids 3rd IARP International Workshop on Humanoid and Human Friendly Robotics, Tsukuba (Japan), 11-12 Dec 2002, pp.83-87.

[13] J. Pettre, T. Simeon, J.P. Laumond. Planning human walk in virtual environments IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'2002), Lausanne, September 2002.

[14] M. Bennewitz, W. Burgard, S. Thrun. Learning Motion Patterns of Persons for Mobile Service Robots. IEEE International Conference on Robotics & Automation (ICRA), 2002.

[15] V. Matellán and R. Simmons, Implementing Human-Acceptable Navigational Behaviour and a Fuzzy Controller for an Autonomous Robot, in Proceedings WAF'2002: 3rd Workshop on Physical Agents, Murcia Spain, March 2002.

Advances in Cogniron

There is no contribution that tackles globally and systematically the problem as we propose to do.

Test Metrics and Incremental Layers:

After 18 months
Design of a motion planner.
After 36 months
A placement planner for various tasks and situations. Perception functions for detecting humans and localizing them relatively to the robot.
After 48 months
Demonstration of the function on a real robot.

CF-MHP: Manipulation in the human presence

Description

The approach is to endow a robot with the ability to assess the feasibility of manipulation tasks it has to achieve in presence and in the vicinity of humans and/or in close coordination with humans, to share the load between the robot and the human and to explain/illustrate (when necessary) a possible course of actions
We will development of algorithms for and executing planning “human-friendly robot manipulation” and, more generally, “human-robot space sharing”.
Besides, robot , task and environment models, the algorithms will deal with human model (motion, manipulation and sensing capabilities) as well as « Social rules » and constraints. Such rules and constraints will be based on the lessons and recommendations obtained after user studies conducted by UH and KTH.
Constraints will be “physical” (geometry, kinematics, dynamics) and “social” (security, acceptability, legibility). An example of a task we plan to demonstrate is a mobile robot equipped with a manipulator that holds an object out to a human or takes it from the human.

State of the art

See CF-NHP

Advances in Cogniron

There is no contribution that tackles globally and systematically the problem as we propose to do.

Test Metrics and Incremental Layers

After 18 months
Design of a manipulation planner.
After 36 months
A planner for various tasks and situation. Perception function for detecting humans and their postures.
After 48 months
Demonstration of the function on a real robot.

CF-TBP: Tracking of human body parts for observation

Description

The tracking of human body parts builds a basis for human robot interaction and for observing a human demonstrator. Based on a 3d human model the cognitive robot companion should be able to detect and track a human and his/her body parts. The tracking system should be robust against lighting changes and fast for real-time processing. Another important aspect of the system is its ability to adapt the model to personal parameters like the size of the limbs of the person. The extracted positions of head, torso, arms and hands can then be used by other cognitive functions in order to recognize the user's action. Different sensors will be used to track human body parts. These are colour vision and stereo colour vision, a time-of-flight camera, laser scanner and microphones. Therefore methods for fusing different sensor data will be investigated.

State-of-the-art

Tracking of body parts has a long history. There exist many approaches for tracking hands and head using skin colour information ([1]). A famous approach for detecting faces in grey level images was introduced by Viola and Jones in [2]. They build a cascade of simple feature detectors resulting in a fast and robust face detector. For the detection and tracking of the full body, there exist different approaches. Ben-Arie et al. [3] use background subtraction of still 2d images to segment the human and then evaluate hypotheses about the position of each body part relative to the torso in order to label each body part. Sidenbladh [4] works on 2d images too. She uses a condensation filter approach to extract the full 3d body pose out of monocular images. In [5] multiple cameras are used to acquire voxel data for the detection and tracking of human body parts. Demirdjian et al. [6] use an ICP-algorithm to fit a 3d human model onto the data of a stereo camera pair.

Literature

[1] M. Störring, T. Kocka, H. J. Andersen, and E. Granum: "Tracking regions of human skin through illumination changes", Pattern Recognition Letters, Special Issue, 24(11), pages 1715-1723, July 2003.
[2] P. Viola, M. Jones: "Rapid Object Detection using a Boosted Cascade of Simple Features", Proc. Of the 2001 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01), Vol. 1, pp. I-511-I-518, Dec. 8-14, 2001, Kauai, Hawaii, USA
[3] J. Ben-Arie, D. Sivalingam and S. Rajaram: "Probabilistic Learning Labeling of Human Body Parts", IASTED International Conference on Circuits, Signals and Systems, (IASTED'03), Cancun, Mexico,pp.275-280 , May 2003.
[4] H. Sidenbladh: "Probabilistic Tracking and Reconstruction of 3D Human Motion in Monocular Video Sequences", PhD Thesis TRITA-NA-0114, ISBN 91-7283-169-3, Dept. of Numerical Analysis and Computer Science, KTH, Sweden 2001.
[5] I. Mikic, M. Trivedi, E. Hunter, P. Cosman: „Human Body Model Acquisition and Tracking Using Voxel Data", Int'l Journal of Computer Vision, 53(3), pp. 199-223, 2003
[6] D. Demirdjian, T. Darrell: "3-D articulated pose tracking for untethered deictic reference" Proc. IEEE Int'l Conf. On Multimodel Interfaces 2002, pp. 267-272, 2002
[7] R. Urtasum, P. Fua: "3D Human Body Tracking using Deterministic Temporal Motion Models", European Conference on Computer Vision, pp. (III)92-106, Prague, Czech Republic, May 2004

Advances in Cogniron

The overall goal of this Cogniron function is to implement a fast and robust tracking of humans and its body parts in 3d. It could be seen as a basic functionality which offers information about the human which could then be used for further processing (e.g. gesture recognition). To achieve this different sensory data will be used and combined.

The advances in this Cogniron function are:

- Definition of precise 3d human model which holds all extracted information about the tracked human. This model also serves as a basis for further activity analysis.

- Design of a framework for multi-modal human body part tracking. The tracking will incorporate different sensory data including stereo vision, colour vision, laser scanners and sound sources.
- Accurate tracking of forearms and hands in 3d depth data without magnetic field tracker or data gloves.
- Robust real-time tracking.
- Combination of colour and depth sensors: Different modalities will be combined in order to enhance tracking results.

Test Metrics and Incremental Layers:
To test the tracking system, several sequences of moving humans will be defined. These allow for testing:

- (Self-) Occlusions
- Different lighting conditions
- Complex poses (e.g. pointing towards the camera, arms along the torso)

The test setup will also contain sequences which are suitable for activity recognition.
Incremental Layers:
After 18 months
A 3d human model will be defined and two different approaches will be implemented for 3d model fitting. One using depth data and the other one working on monocular images. Model adaptation techniques will be investigated.
After 36 months
Advanced fusion methods for combining different sensor sources. Enhancement of the robustness of 3d model fitting.
After 48 months
With increasing complexity the restricted computing power has to be taken into account. Therefore the methods will be enhanced concerning robustness and speed allowing real-time tracking of human body parts.

CF-ACT: Detection and interpretation of human activities and postures

Description

One of the project's aim is that the robot should be able to interact and help humans in household environments. In order to decide whether interaction or help is appreciated, the robot needs to understand the human activity to enable the robot to be able to reason about the user's intention.
An activity model will be set up to represent semantic activities performed by the user. This activity model relies on the information of the human attribute model which describes the user at a geometric level. This includes static poses of the human as well as basic motions of body parts. The activity model adds background knowledge about human activities to achieve a semantic description of human activities.

State-of-the-art

In recent years with improved human tracking systems, the interpretation of human activities became more important for human robot interaction. An activity can be seen as a motion pattern which is temporally periodic and possesses compact spatial structure ([3]). The authors use curvatures of trajectories to classify hand activities.
In [1] human actions are grouped hierarchically as a tree. Each node represents a category of actions. For each category specific features are extracted and a specific HMM is used to classify the action.
Lokman and Kaneko ([2]) use vector displacements of each limb to classify basic actions. These basic actions can then be combined at a higher level to extract higher level actions which incorporate multiple body parts.
One possibility to describe human activities is to use natural language descriptions. Herzog and Rohr [4] use a geometrical scene description to establish the link between the lower level vision system and the higher level scene analysis. At this higher level they are able to describe the movement of pedestrians crossing a street using natural language.
Concept hierarchies of actions are used in [5] to describe human activities. The description of an activity starts at a coarse level and is refined by adding additional information, e.g. "walking" is extended to "move slowly" by adding the type of speed of the movement. The link between image features and natural language description is established by extracting semantic features which correspond to geometric features.

Literature

[1] Taketoshi Moro, Yushi Segawa, Masamichi Shimosa and Tomomasa Sato, "Hierarchical Recognition of Daily Human Actions Based on Continuous Hidden Markov Models" in Proc. of the Sixth IEEE Int'l Conf. On Automatic Face and Gesture Recognition (FGR'04), 2004, Seoul, Korea, May 17-19
[2] Juanda Lokman, Masashide Kaneko, "Hierarchical Interpretation of Composite Human Motion Using Constraints on Angular Pose of Each Body Part" in 13th IEEE Int'l Workshop on Robot and Human Interactive Communication (RO-MAN 2004), 2004, Kurashiki, Okayama, Japan, Sept. 20-22, pp. 335-340
[3] Cen Rao, Mubarak Shah, "View-Invariance in Action Recognition", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01), 2001, Kauai Marriott, Hawaii, Dec. 9-14, pp. (II)316 - (II)322
[4] Gerd Herzog, Karl Rohr, "Integrating Vision and Language: Towards Automatic Description of Human Movements" in Proc. of the 19th Annual German Conference on Artificial Intelligence (KI-95), 1995, Bielefeld, Germany, Sept. 11-13
[5] Atsuhiro Kojima, Takeshi Tamura, Kunio Fukunaga, "Natural Language Description of Human Activites from Video Images Based on Concept Hierarchy of Actions" in International Journal of Computer Vision, 2002, Vol. 2(50), pp. 171-184

Advances in Cogniron

The overall goal of this Cogniron function is to provide a detection and interpretation of human activities. The recognition of these activities is useful for human robot interaction and for teaching the robot.
The advances in this Cogniron function are:

- Definition of a human activity model which is able to represent activities at different levels of detail and which uses the information from the human model developed in CF-TBP
- Detection of human activities based on 3d model data and body part tracking
- Interpretation of human activities

Test metrics

In order to test the detection and interpretation of human activities several sequences of different humans performing different activities will be defined. These will be related to the test sequences of CF-TBP. These allow for testing:

- Detection and Interpretation of activities of a single user
- Detection and Interpretation of the same activity of different users

After 18 months
An activity model will be defined with respect to the human model developed in CF-TBP. Based on this human model, motion primitives will be defined which allow for the detection and interpretation of human activities. Posture classification (still, non-moving postures) will be developed. A detection and interpretation of a first set of human activities will be developed. This set will only include simple activities.
After 36 months
The set of activities will be expanded and will comprise more complex activities. Interaction of human with objects (manipulation, transport, pointing) will be investigated.
After 48 months
The set of activities will comprise typical activities in household environments. This will include combined activities like transporting an object, pointing while walking. Activities between humans will be investigated.

CF-LCT: Learning complex task descriptions

Description

Learning and reasoning are two of the most important capabilities of a Cognitive Robot Companion, since these facilities enable it to show highly flexible, adaptable and humanlike behaviours. These robots should be able to adapt in a flexible and intuitive way to the individual and diversifying environs and the needs of the robot's user. Faced with the need to carry out tasks that can not be anticipated during the robots construction time, a Robot Companion has to learn new tasks. Since the robot is supposed to coexist with humans the learning process can be done through observation of tasks executed by humans or in addition it can be guided by human interaction. This means, the robot has to learn autonomously from user performances and interact with humans for gathering more information, which it can not observe. To understand and analyze demonstrations and to learn complex and larger tasks from observations, it is mandatory for the Companion to be equipped with previous knowledge. The newly acquired task knowledge has to be filed in his memory in a way that it can be easily retrieved and reused both for the reproduction and execution of the performed demonstration and for the analysis of newly observed demonstrations. For the organization of acquired task knowledge the recognition of similarities across different learnt tasks represents a major issue.

State of the art

Several systems and approaches based on programming by demonstration (PbD) have been proposed during the past years. Many of them address special problems or a special subset of objects only. An overview and classification of the approaches can be found in [1], [2]. Basis for the mapping of a demonstration to a robot system are the task representation and task analysis. Often, the analysis of a demonstration takes place observing the changes in the scene, described by using relational expressions or contact relations [3].
Given the whole set of basic operations including the relevant pre- and post-conditions, stated in relational expressions, planning algorithms can be used to generate a sequential task description to reach a defined goal (e.g FF-planner [4]). As most goals allow for different paths, optimization criteria have to be applied to the planning algorithm. These criteria are not necessarily transparent to a human, which can result in the robot performing tasks in an unpredictable or even strange and uncanny way. Integration of task learning methods within a planning system can on the one hand help to solve these problems, and on the other hand enhance system performance by including learned task descriptions as atomic actions into the planning algorithm.

Literature

[1] R. Dillmann, O. Rogalla, M. Ehrenmann, R. Zöllner, and M. Bordegoni, *Learning Robot Behaviour and Skills based on Human Demonstration and Advice: the Machine Learning Paradigm,* in 9th International Symposium of Robotics Research (ISRR 1999), Snowbird, Utah, USA, 9.-12. Oktober 1999, pp. 229–238.

[2] S. Schaal, *Is imitation learning the route to humanoid robots?* in Trends in Cognitive Scienes, vol. 3, 1999, pp. 323–242.

[3] Wang, Q., R. Yang und W. Zhang (2003). *Skill acquisition from human demonstration using FCM clustering of qualitative contact states.* In: Proc. 2003 IEEE Intl. Symp. on Computational Intelligence in Robotics and Automation, July 16-20, 2003 Kobe, Japan., S. 937–942.

[4] Hoffmann, Jörg (2001). *FF: The Fast-Forward Planning System*. The AI Magazine, 22(3):57–62.

Advances in Cogniron

A System, based on background knowledge, will be developed that is able to reconstruct complex tasks, perceive coherences and represent them in an abstract manner in order to use them for further performances, observations and executions. The abstract task description will be used as an input to a task planer, enhancing the capabilities of the system to deal with more general or maybe different tasks

and situations. On the other hand planning capabilities of the robot can be used to optimize the attention of the system in means of directing it to the features that are expected to be most relevant according to the task to be learnt.
During the period of time of the project, this CF will make the following advances:

- Finding a representation and a model for complex tasks, containing coherences between elementary operations.
- Enhancing observation and analysis of task demonstrations based on the developed model and additional background knowledge on the robots own capabilities, in order to extract coherences between the elementary operators.
- Generalization and abstraction of the observation in order to ensure broader reuse possibilities.

Test Metrics and Incremental Layers

To test the learning capabilities for complex actions the domain of household tasks is chosen, and restricted to manipulation of objects and fetch-and-carry-tasks. The focus is on developing a general and scalable framework, which is able to perform the learning of complex tasks.

After 18 months

The framework for learning complex task descriptions from elementary operations will be defined, set up and implemented.
A base of elementary actions according to the household domain will be defined and basic tasks based on them will be learnt.

After 36 months

Complex actions and operations will be learnt and stored in a task database in order to be retrieved and reconsidered for reuse, extension or enhancement.

After 48 months

The system is able to reason autonomously about the learnt tasks, exploit found coherences and dependencies and generalize over the acquired task knowledge.

CF-RG: Learning to reproduce gestures

Description

Learning is a fundamental prerequisite for open-endedness. If a robot companion has to learn to execute a task it can use the way the human executes this task as a hypotheses for its own solution. If the task is a manipulation task then the robot can observe the movement of the demonstrators arm to reproduce the same. The robot can use this information to figure out how to solve the task given. CF-RG addresses the issue. Also communication gestures can be learned by the robot. To solve the reproduction problem the robot has to map the trajectory observed onto its own kinematical setting. This is known as the correspondence problem. For an overview of state of the art and the work on this in the Cogniron project see the forthcoming paper:
Sylvain Calinon, Florent Guenter, Aude Billard, Goal-Directed Imitation in a Humanoid Robot, Submitted to ICRA 05, April 18-22 2005, Barcelona, Spain.

CF-LIF: Learning important features of a task

Description

The trajectory which can be found by CF-RG is does not necessarily contain the information needed to solve a task. Rather, the goal state after execution or other information can be more relevant to learn the task. The function CF-LIF addresses the problem of how the robot can figure out the important hints autonomously. For an overview of state of the art and the work on this in the Cogniron project see the forthcoming paper:
Aris Alissandrakis, Chrystopher Nehaniv, Kerstin Dautenhahn and Joe Saunders, An Approach for Programming Robots by Demonstration to Manipulate Objects : Considerations on Metrics to Achieve Corresponding Effects, Submitted to ICRA 05, April 18-22 2005, Barcelona, Spain.

1.3 Robot Home Tour

Description

Objectives

The key-experiment “Robot Home Tour” stresses the informational human-robot interaction to learn the geometry, topology of the environments and its artefacts, geometry and the identity and location of objects and their spatial-temporal relations. At the current stage it is planned to demonstrate and assess the following abilities which are implemented on a robot platform:

- Basic human-robot dialogue (relates to RA1)
- Recognition of a person and basic (i.e. pointing, waving) gestures (relates to RA1, RA2, RA3)
- Active perception (relates to RA5)
- Acquire geometrical maps of the environment, instantiation of the concept of rooms (relates to RA5)
- Navigation in an un-instrumented home setting (through doors, around dynamic obstacles,…) (RA5)
- Autonomous Decision for control and coordination of own functions (RA6)

Script

We constructed a dining room and the user is named Tom. In the first stage the user shows the robot the dining room in counter clockwise direction; in the second stage, we assume that the robot already acquired some knowledge about the objects in the dining room in the first stage. Having successfully completed the home tour, the robot can now initiate some more intelligent dialogs.
A typical storyline for the key-experiment looks like this:

1. The robot looks for human communication partners in an office like environment.
2. The robots finds a person and directs its attention to this person. If this person "registers" itself by greeting the robot (e.g. 'hello'), the robot focusses its attention on this partner and is not distracted by other humans.
3. The partner asks the robot to follow the human and the robot complies.
4. The partner points to various objects and tells the robots their names e.g. plant and coffee cup.
5. The partner tells the robot the name of its the current location, the kitchen.
6. The partner asks the robot to follow him/her again and the robot complies.
7. The partner again tells the robot the name of its new current location, the living-room.
8. The partner asks the robot to go to the kitchen and the robot complies.
9. If the robot cannot find a route to the kitchen (i.e., all routes are blocked) the robot will ask for assistance. Otherwise it drives to the kitchen.
10. As the robot arrives in the kitchen another partner asks the robot to search for the coffee cup.

The robot drives to the last known location of the coffee cup and reports finding it or not finding it.

Implementation

Robot

Hardware

The hardware of this key-experiment will be based on the BIRON robot at UBI. Additional sensors and computing hardware may be added in order to realize the CFs coming from other partners. The following Table gives an overview of the components which will probably used.

Component	Description
Robot	Camera height: 1.425 m Microphone height: 1.160 m Laser range finder height: 0.3 m;
Camera (Pan-Tilt)	(Photos that are taken by it are used for persons, objects and gesture recognition.) Type: Sony EVI-D31 (with Pan-Tilt-Unit) Resolution: 768 * 576 Pixels (PAL) Adjustable features: zoom, pan-tilt, focus angle of view: 48.8°
Camera (stereo)	(Photos that are taken by it are used for persons and gesture recognition.) Type: Videre Design Resolution: 640 * 480 Pixels Adjustable features: none angle of view: 90°
Touchscreen	Size: 12“;
Microphones	Type: AKG Acoustics C400 PC Computer Microphone (We did some changes on the preamplifier of the microphones to improve their performance) Speech signal saving Sampling frequency: 16 kHz Sampling resolution: 16 bit/sample (signed)
Preamplifier	Vivanco MA 222

Fig. 0-1: Hardware components in the Robot Home Tour

Software

Operating System: Linux
Programming Language: C/C++
Communication Mechanisms: XML, TCP-IP

Environment

Physical

Technical constraints that need to be fulfilled during interacting with the robot in the subsequent script:

- Minimal distance between a target object and the robot: 1.5m;
- Minimal distance between the user gesture and the robot: 1.5m;

- Maximal distance between the user and the robot: 3 m;

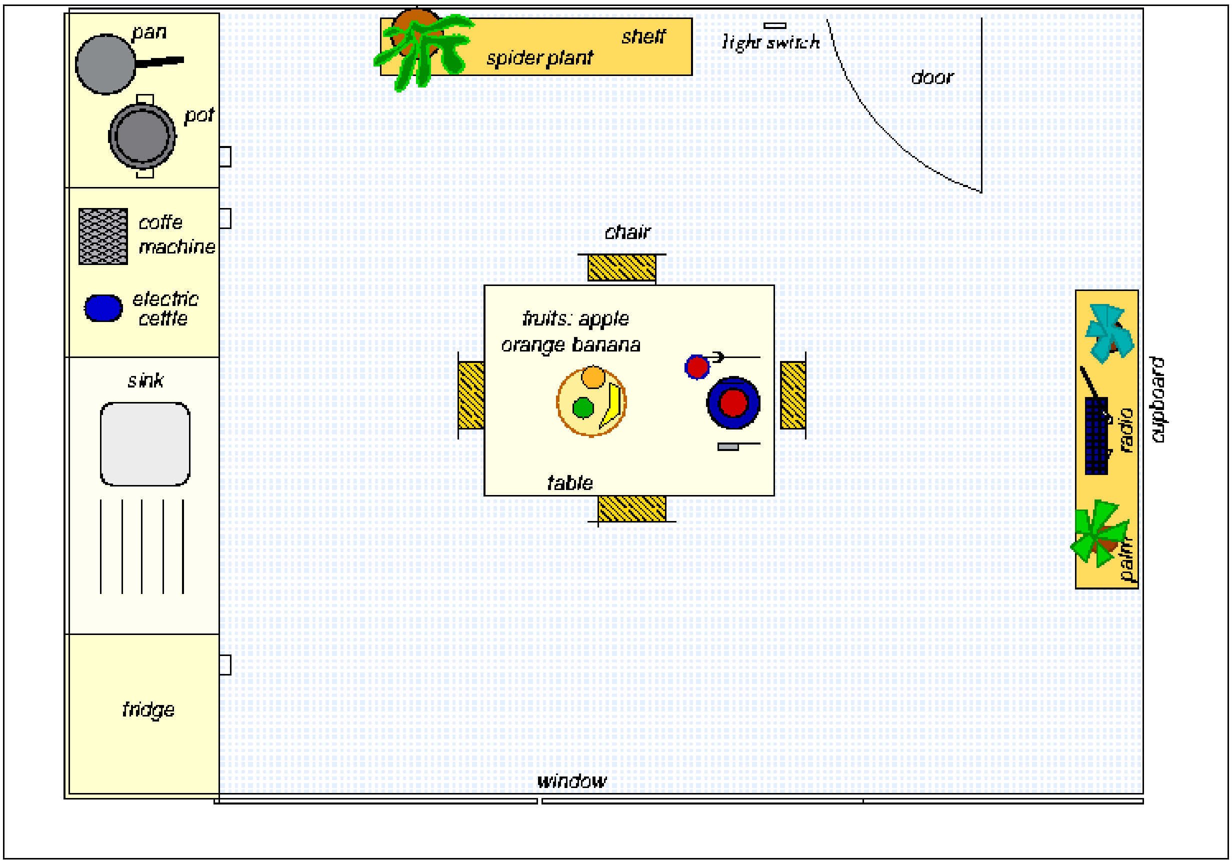

Fig. 0-2: A typical room for the Robot Home Tour

Static Objects
Objects: shelf, table, chair, sink, cooker, fridge and wall with light switch. Objects should be put in normal daylight.
Movable Objects

Cupboard	Desk	Dining table	Kitchen Items
book (standing)	telephone	spoon	coffee machine
plant	book (isolated)	knife	electric kettle
mug	keyboard	fork	pot
boxes	monitor	plate	pan
tea caddy	lamp	cup	teapot
radio	mouse	glass	
	puncher		

Fig. 0-3: Movable objects relevant to the Robot Home Tour

Animate Objects
Humans (for person tracking).

Social

Interaction Style	Interactive	User, talking to robot
	Non Interactive	Bystander, talking to someone else
Familiarity	Expert	The user knows how to interact with the system and gives correct instructions
	Naive user	This user may occasionally violate the constraints of the system. This should result in an appropriate reaction of the robot, but it may also result in failures.

Fig. 0-4: Classification of users in the Robot Home Tour

1.4 The Curious Robot

Description

Objectives

The key-experiment “Curious Robot” stresses the perceptional skills of the robot to acquire knowledge about the locations and properties of objects and their spatial-temporal relations. At the current stage it is planned to demonstrate and assess the following skills which are implemented on a robot platform.

- Recognition of a limited number of known (learned) objects (RA5)
- Locate and identify a person (RA2, RA3)
- Collaboration with human (RA3, RA6)

Script

We want this scenario to really be a framework for research on cognitive capacities, and more precisely "learning", "taking initiative", "intentionality attribution and manifestation" and "curiosity", as the manifestation of cognition. A set-up containing a room, a table, a person, several objects on the table will present it. The robot will be a mobile manipulator.

Script 1: Robot initiative

In this set-up the robot anticipates a situation that may occur and acts in order to facilitate the future action of the person.

A person is sitting and busy with a task. The robot is doing its own tasks and also observes human who would be doing some gestures, not meant towards the robot, which usually express a need for a drink (for example, the human "plays"' with an empty cup). The robot interrupts its own task, approaches the human and asks her if she wants a drink. Upon a positive answer, the robot fetches it.

Variation 1: the robot does not ask. It fetches the drink and serves it directly.

Variation 2: there is already a can on the table, but it is too far to be reached by the person. The robot takes the initiative to place it closer to the person.

Here, the main issues are on reasoning about the task, interpretation of human activities, intentionality, and initiative. Another key aspect is linked to the acceptability and to the legibility of robot activities. The robot behaves, performs its task in a way that permits the persons in the room to easily guess/interpret correctly its intentions and to easily interact with the robot.

The robot may be already performing an activity to achieve a goal explicitly requested by the human, or a routine task. The robot will have to balance between all its current and future activities and may decide to merge its different activities.

Script 2: Robot curiosity

Simple set-up (version zero): The robot observes a table top, doing nothing. A person puts an object (say a cup, a pen or a box) on the table. The robot starts sensing the object with all its possible sensors, including grasping it, looking at it from different viewpoints. The robot then ask the human about some properties of the object (what is it? Is it related to other known objects or concepts?)

Here the issues are concept formation, categorization of objects, acquisition of skills to manipulate objects, communication about objects.

Manipulation is only a modality for acquiring knowledge. But it is a necessary one.

Version one: The robot is performing a task, and a new object appears in its vicinity (role of context in concept formation)

Version two: Several new objects; here there is a complexity issue.

In all these variations, we want to investigate the robot as driven by the need to increase its knowledge level or information quantity. This can be theorized. Information can be measured by entropy and considered as a "value" driving robot behaviour. The robot general behaviour being dictated by the drive to increase information. How does this interact with other values and other motivations?

Implementation

Robot

Hardware *(see robot description in D 7.3.1)*

Component	**Description**
Neobotix platform (LAAS)	PA-10 Mitsubishi manipulator, Sick scanner, Stereo camera head (Pan/Tilt-unit) and Screen 
B21 I-robot (LAAS)	Sony EVI-D31, Sick, Micropix Videocamera Firewire, Touchscreen, Soundcard + speakers 

Fig. 0-5: Hardware components in The Curious Robot

Software

Operating system: Linux
Programming Language: A set of tools for developing and integrating functional modules (GeNom) and robot supervision
Communication Mechanisms: TCP-IP

Environment

Physical

Static Objects: Table, Board
Movable Objects: Chair, Can, Marker.

Animate Objects: Human

Social

The Human does not necessarily have the willingness to interact with the robot.

1.5 Learning Skills and Tasks

Description

Objectives

The key-experiment "Learning Skills and Tasks" stresses the learning and reasoning capabilities for the robot to acquire knowledge about goals and tasks employing the example of laying out the table. At the current stage it is planned to demonstrate and assess the following skills which are implemented on a robot platform:

- Learning goals from observations (RA4, RA5, RA6)
- Reproduction of the goal for arbitrary starting conditions (RA4, RA6)

Script

Script 1: Learning Skills "Arranging and interacting with objects"
The script stresses the robot's ability to learn from implicit (imitation learning) and explicit (verbal interaction) teaching. It learns new skills to manipulate objects and, by so doing, it learns a new task, that of laying out the table. The scenario goes like this:
The robot watches a human demonstrator laying out a cover on a table, i.e. placing plate, cup, fork, knife and napkin. The demonstration is repeated several times. Each demonstration is slightly different from the others. For instance, in one demonstration, the plate is placed first, then comes the glass, fork, knife and napkin. In another demonstration, the fork and knifes are placed first, then, the plate and glass and napkin. In one case, the napkin is folded as a triangle and placed on the plate. In another case, it is folded on itself and placed in the glass. The absolute position of the cover on the table varies as well.
While watching the demonstrations, the robot learns the invariants of the task (relative position and orientation of the objects with respect to one another) and new skills such as object-actions relations (how to grip the cup by its handle).
Once the demonstrations are finished, the robot tries to reproduce the task. While doing so, the robot might query the user if some demonstration were ambiguous and its choice is non deterministic. For instance, it might ask "should I put the napkin in the plate or on the glass"? The user might stop and correct verbally the robot during the reproduction, if the robot makes important mistakes. For instance, if the robot places the fork at the place normally occupied by the knife, the user might tell the robot "this is incorrect", leaving the robot work out which of its actions were incorrect; alternatively, the user might say "no, the fork should go on the other side of the plate", assuming that the robot.
Script 2: Learning Tasks "Serving a guest"
The scenario focuses on the set up of the living room table for a nice evening reading a book or watching TV.
The set-up involves the following objects: a cup coffee, a plate, a box of biscuits, a bottle of water, a glass, a book and maybe other objects which can disturb the task execution of the robot. The robot is equipped with at least one manipulator arm and is mobile. There are four pieces of furniture: a couch where the human sits at, a table which has to be laid out, a book-shelf containing books, and a cupboard which contains dishes.
The first part of the script deals with the teach-in functions of the robot.
Demonstration aspects:

a) The table is free

b) The human asks the robot to follow it to the table

c) The human shows the robot how to put one piece (the cup) onto the table

d) The human shows the robot how to lay out a complex pattern of objects relative to each other

e) The human removes all objects involved and places them onto their dedicated positions in the other furniture

Reproduction aspects:

a) The robot is to produce the setting demonstrated by the human

b) If there are conditions which hinder task execution the robot detect (e.g. a book laying on the target area) the robots detects it and takes suitable measures autonomously

Implementation

Robot

Hardware

Component	**Description**
Robox experimental platform	Wheel-based platform with a 5 DOF arm - 1 DOF gripper, provided with a binocular head (2 CCD camera on a 6 DOFs head)  
Care-O-bot	Wheel-based platform, 6 DOF arm, laser scanners and colour cameras 

Fig. 0-6: Hardware components in Learning Skills and Tasks

Software

Operating System: Windows, Linux.
Programming Language: C/C++
Communication mechanism: TCP-IP

Environment

Physical

The set-up planned involves the following objects: a cup coffee, a plate, a box of biscuits, a bottle of water, a glass, a book and maybe other objects which can disturb the task execution of the robot. The robot is equipped with at least one manipulator arm and is mobile. There are four pieces of furniture: a couch where the human sits at, a table which has to be laid out, a book-shelf containing books, and a cupboard which contains dishes.

Fig. 0-7: Example of a typical environment for Skill and Task Learning

Social

The human plays two roles in this key-experiment. First the user is a demonstrator of a task. Then he/she is the client who gets served.

1.6 Interaction Diagram

The following figure presents the current interaction diagram for the three key-experiments.

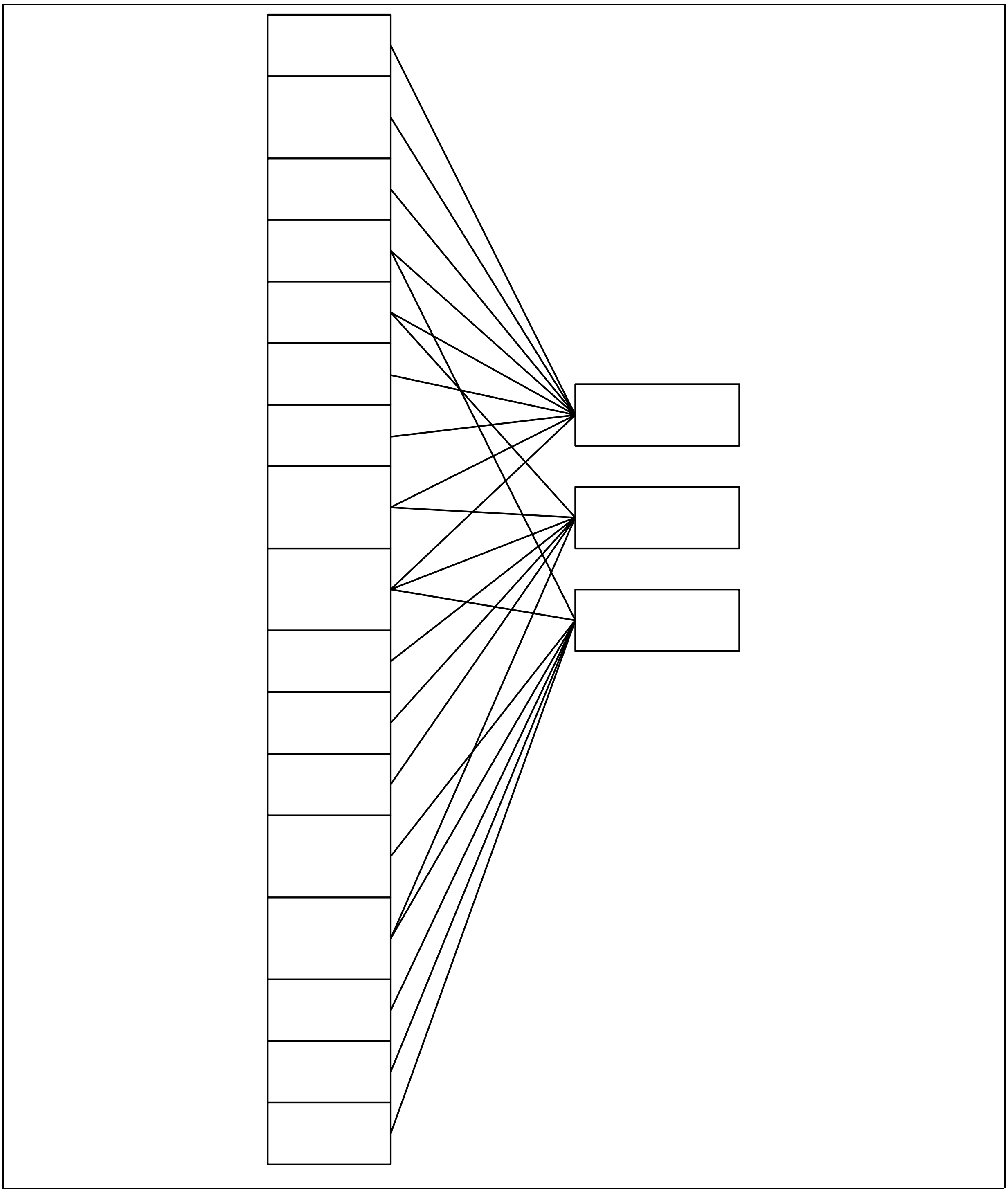

Fig. 0-8: Interaction diagram for the key-experiments with functional decomposition

2 Conclusions and Future work

It turned out to be much harder than expected to get clear and precise definitions of the key-experiments. If the scrips are too specific then a straightforward solution would show the desired behaviour of the robot. Such a solution has not necessarily to be very general. Hence, the definitions have to be coarse to capture the contents of the research areas which are phrased in an abstract scientific manner. To tackle this, functional descriptions were introduced and presented in a consistent way. This allows for at least partial specific definitions. It was very time consuming to collect all the functional descriptions and to relate them to the key-experiments. But it is our hope that this two-way approach of both, top-down and bottom-up analysis results in system descriptions which are suited to reach the goals of RA7. This document certainly is a good basis for RA7 but work has to be continued.

The next step in this work package will be the refinement and update of the specification document in terms of:

- Clarification and detailed planning of partner interactions on the basis of the functions defined in the first phase
- Further detailed functional descriptions (input/output data) and classification of the functions defined in the first phase
- Aspects on how the functions defined in the first phase can be integrated into a key-experiment on hard- and software level
- Test schemes of how the functions can be tested in the framework of the key-experiments
- Methods on how can a key-experiment be evaluated as a whole

3 References

3.1 Applicable documents

The deliverables D7.2.1, D7.3.1 and D7.4.1 are directly related to this deliverable.