Towards reorientation with a humanoid robot

Dietmar Bruckner[1], Markus Vincze[2], Isabella Hinterleitner[1]

[1] Institute of Computer Technology, Vienna University of Technology,
Gusshausstrasse 27/384, 1040 Vienna, Austria, Europe
{bruckner, hinterleitner]@ict.tuwien.ac.at
[2] Automation and Control Institute, Vienna University of Technology,
Gusshausstrasse 29/384, 1040 Vienna, Austria, Europe
vincze@acin.tuwien.ac.at

Abstract. Current cognitive vision systems and autonomous robots are not able to flexibly adapt to novel scenes. For example, when entering a new kitchen it is relatively simple for humans to adapt to the new situation. However, there exist no methods such that a robot holds a generic kitchen model that is then adapted to the new situation. We tackle this by developing a hierarchical ontology system linking process, object, and abstract world knowledge via connecting ontologies, all defined in a formal description language.

The items and objects and their affordances in the object ontology are either learned from 3D models of the Web or from samples. We bind the features of the learned models to the concepts represented in the ontology. This enables us to actively search for objects to be expected to be seen in a kitchen scenario. The search for the objects will use the selection of cues appropriate to the relevant object. We plan to evaluate this model in three different kitchens with a mobile robot with an arm and further with Romeo, a humanoid robot designed by Aldebaran to operate in homes.

Keywords: Cognitive Robotics, Situated Vision, Ontology, Reorientation

1 Introduction

Our goal is to understand the necessary preconditions for perception and action in a context which is familiar in an *abstract* sense (e.g., tea making), yet not in a given *concrete* setting (a kitchen not seen before). Our vision is to give an explanation of this interplay between perception and vision and to provide a representation of these different types of knowledge, such as different ontologies [1] (encompassing task, process, affordance [2] and object knowledge) that are used at different stages. This abstract knowledge needs to be modeled such that it can be tethered [3] to actual objects and the affordances provided by a new situation. We see several major advantages of this approach:

- Clean separation of top-down abstract knowledge and bottom-up concrete data.
- Insights of what an abstract database needs to provide in order to guide visual input towards the task affordances and processes (the actual robot grasping and manipulation).

- A thorough understanding of what vision needs to provide to perceive these affordances given a specific task and robot embodiment, we refer to this as the *situatedness* of vision [4].
- The natural inclusion of *attention-based* vision to re-orientate in novel environments extending present artificial saliency maps to situated robots.
- And, formal means to study how to flexibly adapt to novel settings given the genericness of the ontological knowledge.

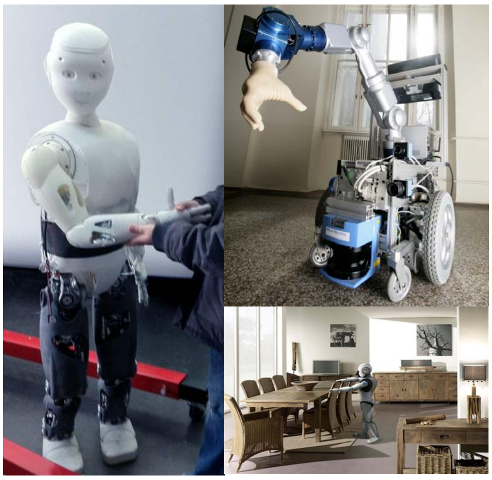

Fig. 1. Robots to show the situated vision approach: James with 7DOF arm and hand prosthesis (top right) and Romeo by Aldebaran (140cm tall, right bottom) shaking hands with a boy (left).

2 Motivation

In his seminal work Michael Land investigated human eye movement when performing a tea-making task in a kitchen at their university [5]. The important aspect is that the subjects had not seen the kitchen before, but before recording the sequence it took them about 30-40 seconds to re-orient in the new setting. During the recorded eye fixations while making tea all fixations are "relevant" to the task, clearly indicating that the human has built up a full mental representation of the kitchen setting. The fixations fall into four action types: locating (objects), directing (an object towards a new location), guiding (the interaction of one object with another), or checking (the state of a variable). The authors outline different levels of description, starting from a higher-level goal ("make the tea") to single steps (such as "transport to sink").

It is striking that this work focuses on questions involved in the "execution" of the task and not the crucial re-orientation phase before the task can be executed so smoothly. We are, however, interested in the seconds *before* the test subject starts making the tea, i.e., the seconds that involve the "binding" of needed parts of the task to the concrete objects found in the current situation [6].

3 State of the Art

3.1 Situated Vision

AI's notion of "situatedness" [7, 4] emphasizes the necessity to focus the operation of an agent enabling it to operate reliably in an uncertain, changing environment. The rationale is that domain knowledge scales the possibilities found in such a setting.

Accomplishing tasks in an everyday setting, especially in a home environment, is a great challenge for robots and vision alike. The reason is the inherent uncertainty that is imposed by the lack of controlling the environmental conditions. Humanoids or rolling torsos can approach an immediate target but environments are fully modeled and objects need to be clearly colored or textured [8, 9]. The environment can be reconstructed from tracking features using a Manhattan assumption [10], however obtaining high level structure requires constraints such as given models [11] or recurrent items [12]. First work shows how to systematically search for objects using bottom up information in a room with tables and sofas [13], a few shelves with a humanoid robot [14], or learning views related to a bicycle [15]. The most cognitive approach learns and reasons about object to room relations [16], which is a first step towards reorienting in a new setting.

3.2 Formalizing Task Knowledge

Task planning is a key issue in robotics. [17] defines a task matrix and focuses on developing a set of complex behaviors using manually-devised connections between task programs. [18] studies industrial automation tasks and distributes tasks amongst agents, as robots in a flexible factory production have to solve industrial automation tasks. The complete ontology of how to achieve a task is split up in the sum of partial task descriptions in the agents. A theoretical approach for an ontology of robot tasks is presented in [19].

While psychology is debating the necessity and existence of explicitly accessible representations [20] we have to find ways to assign tasks to robots to make them interact with the environment. The robot has to check the environment for available resources. In our situated vision approach this is achieved via features or objects in the environment that have to be recognized in terms of *affordances* [2]. Gibson defines an affordance as follows:

Affordances relate the utility of things, events, and places to the needs of animals and their actions in fulfilling them [...] Affordances themselves are perceived and, in fact, are the essence of what we perceive.

3.3 Visual Attention

The concept of visual attention is well investigated in human vision. Many psychological models of visual attention exist (cf. overviews in [21, 22]). Among the best known models are the Feature Integration Theory [23] and the Guided Search model [24]. All models have in common that several feature channels are processed in parallel and fused to a single saliency map. This map highlights the most interesting regions in a scene and guides the focus of attention. There are controversies about which features guide the deployment of attention. Some cues however are undoubted to belong to these basic features: color, motion, orientation and size. Other studies also support depth cues, shape and curvature [25].

4 ENTER approach

We introduce a concept of four ontologies, hierarchically ranging from high level tasks to low level objects. Although the planned nested ontology has to fulfill a high level task, at the level of the task ontology only the object class is defined.

In the first step a task is broken into subtasks (processes). Objects have to be identified to meet requirements of a process (affordances). Thus, on the lower level of ontologies we have a basic object ontology and the affordance ontology.

The fact that only the class is defined at task level makes the whole ontology very much light weighted and flexible.

The robot evaluates the environment according to affordances of, e.g., water supply, making hot water, containers for tea bags and the like. To do so the robot has to identify the function of an object (where the term “object” also includes, e.g., flat surfaces that can be used to place a cup or saucer). Affordances are more than descriptions of objective functionality. Rather than describing the function of an object they describe what impact an object has on the entity that interacts with the object. Thus, it depends not only on the object, but also on the agent and on what the environment can offer. Depending on the shape and degrees of freedom of a robot hand an object may be lift-able or even carry-able or it may not be movable at all and thus considered as an obstacle. Depending on the size of the robot hand an object may be considered as graspable or not. Below object level an affordance-based ontology is constructed: it links objects to robot actions that could be performed to complete a task given the robot's capabilities.

The task ontology is considered as a domain ontology and on the highest level it contains the processes required to fulfill a task including its context.

Finally, the object ontology contains all objects that are needed for a task, such as cup, kettle, milk, fridge, tap, etc. The object ontology subsumes the objects’ affordances representing the semantic knowledge.

The example in Fig.2 shows how the FillCup scenario can be fulfilled based on the model of hierarchical information processing. First, a camera gives information, whether the kettle and/or the cup are already on the table. Then, giving direct feedback during execution of the process, sensors, such as the contact sensors, detect, whether the robot is grasping the kettle correctly. Finally, the sensors observe, if the arm is manipulated in order to pour the water from the kettle in the cup. The example for the robot pouring the water works given the fact that the cup is already in the range of the machine. If this is not the case, the world knowledge will be consulted in order to search where a cup could be located. As is stated there, the robot would then start searching in cabinets for a cup.

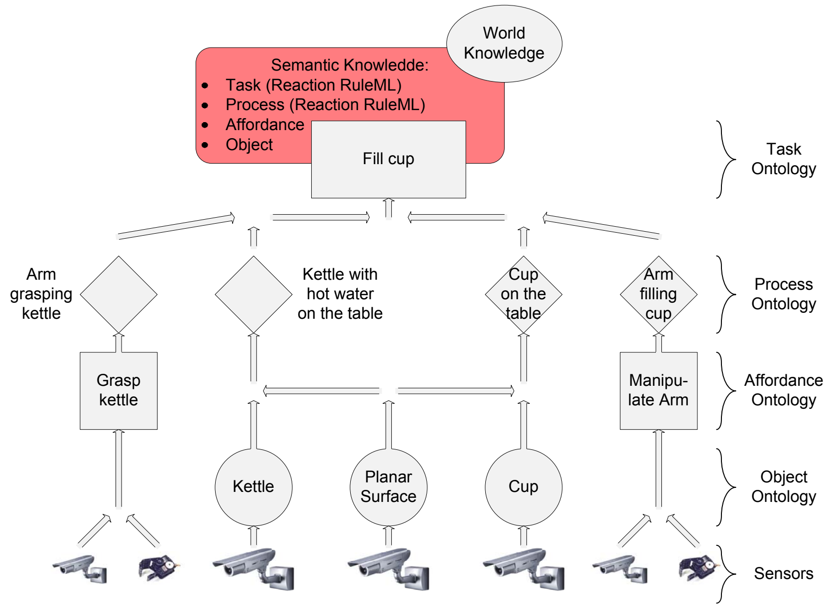

Fig. 2. ENTER Ontologies for the concrete example of filling the tea cup with hot water from the kettle. On the right the different ontologies are listed while on the left the dependencies between instances of the ontologies are shown.

The novelty is that this neuro-symbolic network model can be described by using rule systems, where each network node becomes a rule, representing the node inputs as premises and the node outputs as conclusions. This enables a high-level specification of neuro-symbolic networks and the use of the RuleML format for network interchange [26].

References

1. N. Shadbolt, W. Hall, and T. Berners-Lee. The semantic web revisited. IEEE Intelligent Systems, 21:96-101, 2006.
2. J.J. Gibson. The ecological approach to visual perception. Houghton Mifflin, USA, 1979.
3. A. Sloman. Getting meaning off the ground: Symbol-grounding vs symbol-tethering. Online at: http://www.cs.bham.ac.uk/research/projects/cogaff/talks/ (04/2009), March 2002.
4. R. Pfeifer and C. Scheier. Understanding Intelligence. MIT Press, Boston, 2001.
5. M. Land, N. Mennie, and J. Rusted. The roles of vision and eye movements in the control of activities of daily living. Perception, 28(11):1311-1328, 1999.
6. B.W. Tatler, M. Hayhoe, Michael F. Land, and D.H. Ballard. Eye guidance in natural vision: Reinterpreting salience. Journal of Vision, 11(5):1-23, 2011.
7. W.J. Clancy. Situated Cognition. Cambridge University Press, Cambridge, 1997.

8. F. Gravot, A. Haneda, K. Okada, and M. Inaba. Cooking for humanoid robot, a task that needs symbolic and geometric reasoning. In Robotics and Automation, 2006. ICRA 2006 IEEE International Conference on, pages 462-467, May 2006.
9. M. Tenorth, U. Klank, D. Pangercic, and M. Beetz. Web-enabled robots. Robotics Automation Magazine, IEEE, 18(2):58 -68, June 2011.
10. Alex Flint, Christopher Mei, David Murray, and Ian Reid. A dynamic programming approach to reconstructing building interiors. In Proc. 13th IEEE European Conference on Computer Vision, pages 394-407. Springer, 2010.
11. R.B. Rusu, Z.C. Marton, N. Blodow, M.E. Dolha, and M. Beetz. Functional object mapping of kitchen environments. In Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, pages 3525-3532, Sept. 2008.
12. M. Ruhnke, B. Steder, G. Grisetti, and W. Burgard. Unsupervised learning of compact 3d models based on the detection of recurrent structures. In Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, pages 2137 -2142, oct. 2010.
13. David Meger, Per-Erik Forssn, Kevin Lai, Scott Helmer, Sancho McCann, Tristram Southey, Matthew Baumann, James J. Little, and David G. Lowe. Curious george: An attentive semantic robot. Robotics and Autonomous Systems, 56(6):503 - 511, 2008. From Sensors to Human Spatial Concepts.
14. A. Andreopoulos, S. Hasler, H. Wersing, H. Janssen, J.K. Tsotsos, and E. Korner. Active 3d object localization using a humanoid robot. Robotics, IEEE Trans., 27(1):47 -64, 2011.
15. D. Meger, A. Gupta, and J.J. Little. Viewpoint detection models for sequential embodied object category recognition. In Robotics and Automation (ICRA), 2010 IEEE International Conference on, pages 5055 -5061, may 2010.
16. J.L. Wyatt, A. Aydemir, M. Brenner, M. Hanheide, N. Hawes, P. Jensfelt, M. Kristan, G.M. Kruijff, P. Lison, A. Pronobis, K. Sjöö, A. Vrecko, H. Zender, M. Zillich, and D. Skocaj. Self-understanding and self-extension: A systems and representational approach. Autonomous Mental Development, IEEE Transactions on, 2(4):282 -303, Dec. 2010.
17. Evan Drumwright. The task matrix: An extensible framework for creating versatile humanoid robots. In IEEE Intl. Conf. on Robotics and Automation (ICRA), 2006.
18. Quibin Feng, A. Bratukhin, A. Treytl, and T. Sauter. A exible multi-agent system architecture for plant automation. In 5th IEEE International Conference on Industrial Informatics, pages 1047 - 1052, 2007.
19. S.S. Hidayat, B.K. Kim, and K. Ohba. Learning affordance for semantic robots using ontology approach. In IEEE/RSJ International Conference on Intelligent Robots and Systems IROS, pages 2630 - 2636, 2008.
20. P. Haselager, A. de Groot, and H. van Rappard. Representationalism vs. antirepresentationalism: a debate for the sake of appearance. Philosophical Psychology, 16:5-23, 2003.
21. C. Bundesen and T. Habekost. Attention. In K. Lamberts and R. Goldstone, editors, Handbook of Cognition. Sage Publications, London, 2005.
22. S. Frintrop, E. Rome, and H.I. Christensen. Computational visual attention systems and their cognitive foundation: A survey. ACM Trans. Applied Perception, 7(1):1-46, 2010.
23. A.M. Treisman and G. Gelade. A feature integration theory of attention. Cognitive Psychology, 12:97-136, 1980.
24. J.M. Wolfe. Visual search. In H. Pashler, ed., Attention, p. 13-74. Psychology Press, Hove, U.K., 1998.
25. J.M. Wolfe and T.S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience, 5:1-7, 2004.
26. H. Boley. Posl: An integrated positional-slotted language for semantic web knowledge, ruleml draft. Online at: http://www.ruleml.org/submission/ruleml-shortation.html, 2004.