Journal of Ambient Intelligence and Smart Environments 1 (2009) 311–321
DOI 10.3233/AIS-2009-0040
IOS Press

Grounding commonsense knowledge in intelligent systems

Marios Daoutis *, Silvia Coradeshi and Amy Loutfi
Applied Autonomous Sensor Systems Research Center – Mobile Robotics Lab, University of Örebro, Sweden
E-mail: {marios.daoutis, silvia.coradeschi, amy.loutfi}@oru.se

Abstract. Ambient environments which integrate a number of sensing devices and actuators intended for use by human users need to be able to express knowledge about objects, their functions and their properties to assist in the performance of everyday tasks. For this to occur perceptual data must be grounded to symbolic information that in its turn can be used in the communication with the human. For symbolic information to be meaningful it should be part of a rich knowledge base that includes an ontology of concepts and common sense. In this work we present an integration between ResearchCyc and an anchoring framework that mediates the connection between the perceptual information in an intelligent home environment and the reasoning system. Through simple dialogues we validate how objects placed in the home environment are grounded by a network of sensors and made available to a larger KB where reasoning is exploited. This first integration work is a step towards integrating the richness of a KRR system developed over many years in isolation, with a physically embedded intelligent system.

Keywords: Physical symbol grounding, commonsense knowledge representation, human robot interaction, intelligent home

1. Introduction

In the past years we have seen a great emphasis on research that focuses on robotic assistants and smart homes helping users to perform more and more complex tasks and everyday activities. We can see emerging paradigms where robotic devices, simple sensors pervasively embedded in everyday environments and humans communicate and cooperate. Such paradigms include ubiquitous robotic systems, symbiotic systems, intelligent spaces and robot ecologies and combine visions from autonomous robotics and ambient intelligence to integrate both simple and complex devices. The next step in embedding those devices into our everyday lives, will require that they can be operated by non-experts that have no knowledge about their internal functioning. Thus, the key problem lies in enabling the interaction between such systems and the human. One intuitive solution is communication on the conceptual level, through natural language. With this kind of communication we meet two secondary problems: establishing the required amount of background knowledge in order to successfully enable communication through natural language, and how to ground perceptual information into background knowledge, referred to, as "Symbol Grounding Problem" [23]. If a smart home or robotic assistant will employ a successful natural language dialogue it will certainly use concepts to refer to things and experiences in the world and these concepts must be grounded to perceptual data that refer to these objects. Since our intelligent systems are dedicated to operate and communicate in the real world, they need access to general commonsense knowledge. In this paper we address a framework which deploys a knowledge representation and reasoning system within a symbiotic system. A key feature of the system is the integration of the KRR together with information coming from physical perceptual data originating from a real smart home which includes a robotic assistant. This integration facilitates the mapping of non semantically structured perceptual information into pieces of knowledge that conform with a shared vocabulary (ontology). It allows for the information to be hierarchically structured, defin-

*Corresponding author.

ing concepts at different levels of abstraction. Furthermore, information can be exchanged and queried between agents including humans. We focus mostly, but not only, on the iconic representation [23]. While communication between artificial agents on the perceptual level has been studied in the recent past [24], we are going to study communication on the conceptual level which can be used by both artificial agents and humans simultaneously. In past research we have seen attempts to model or integrate commonsense knowledge to mobile robots, which were theoretical [13] or addressed a small subset of commonsense information in real world scenarios [14]. Recently we see improvements as this subject is brought back to life for the purpose of Visual Scene Detection and Interpretation [12] and practical and grounded knowledge representation systems using commonsense information [15].

Generally the integration of KR&R with robotic systems has been an increasingly interesting topic for cognitive robotics. Examples are found in semantic mapping [1], improving planning and control aspects [2], and most notably HRI systems [3]. Typically, such systems have used small KR&R systems, tailored to the specific application at hand. We investigate the possibility to include a fully fledged KR&R system that has been developed independently of the intended application and has been maintained over an extended period of time. Finally a common observation for systems dealing with symbol grounding is that the majority of works disregard a generic solution to physical symbol grounding by hard coding ad hoc solutions or limiting the domain to a very small subset (i.e. addressing only the spatial relations, or only the topological localization etc.). In this work since we are interested in integrating a KRR system with the perceptual data lying in a symbiotic system which may include a number of heterogeneous sensors and complex networks of sensors (i.e. a robotic system) we use an anchoring framework, which creates and maintains in time the connection between the symbolic system and the perceptual data corresponding to physical objects. A further advantage exploited in this work is the use of common sense reasoning which is available in the KR&R components. In addition to enabling queries from the human users this common sense knowledge is used to reason about objects, their properties and functionalities.

The overall system is evaluated in a testbed consisting of a number of sensors, a mobile robot and a human user. It should be emphasized that the presented work focuses on the system integration and is built upon a number of existing components which have been tested previously in isolation. Thus the key novelty is the integration of the components and the validation of the complete system in a real home-like environment.

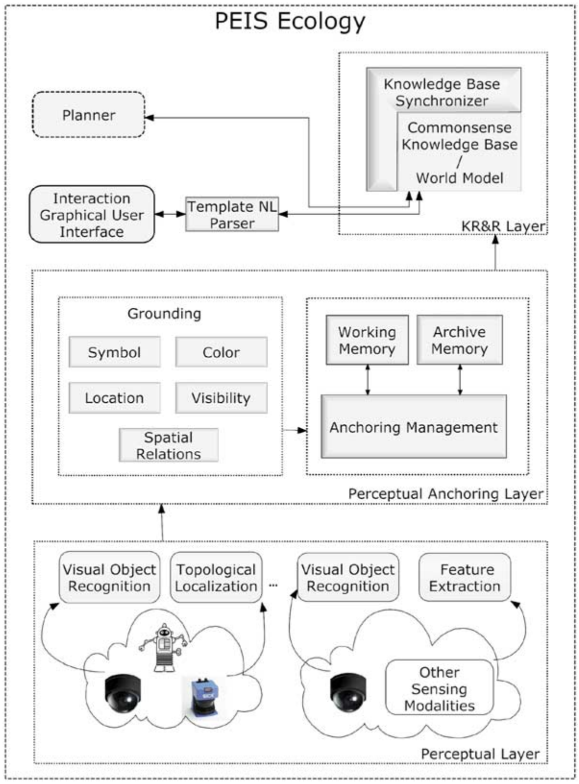

Fig. 1. Overview of the implemented robotic system for knowledge based perceptual anchoring. Arrows indicate the flow of information.

2. System overview

The implemented architecture lies within our realization of a smart environment called PEIS- Home (PEIS stands for Physically Embedded Intelligent Systems) which is described in Section 3. The system is divided in three main layers of computation as depicted in Fig. 1. Namely there is a perceptual layer, a symbol grounding and maintenance layer, and finally the Knowledge Representation and Reasoning Layer. Each layer can be conceived as a separate cognitive ability of the robotic assistant or smart home. In the perceptual layer, data originating from hardware sensors are handled as groups, and are filtered, according to their complexity, by appropriate feature extraction and matching algorithms. For example, one perceptual group includes the sensors (camera, rangefinder, sonars, etc.) that are mounted on the mobile robot, while another perceptual group concerns the sensors that are located in each room of the smart home (person recognition / tracking, temperature / luminosity / humidity sensors, cameras etc.). Information from all the perceptual groups, is then fed into the symbol grounding layer. The mechanism we adopt for symbol grounding is based on perceptual anchoring [8,9] which is responsible for three functions. It grounds each piece of perceptual information into its symbolic counterpart, while then it encodes the symbolic translations into a data structure that denotes a physical entity. Finally it manages all those encodings in time. In the knowledge representation layer which contains a commonsense knowledge base (KB) and inference mechanisms, all the grounded information from the symbol grounding layer, is instantiated into concepts from the KB. Finally this information will be used, through a simple NL system to facilitate a dialogue with the human user.

2.1. *Perception*

The perceptual layers consist of a number of heterogeneous sensors organized in perceptual groups which publish perceptual information to the PEIS communication middleware and are used for further processing. As mentioned in the introduction we heavily rely on visual processing (iconic representation) to be able to extract later on semantically rich information about the environment. For visual processing, the experiments presented adopts a SIFT-based approach [5]. We developed an algorithm which processes the image feeds from any video source lying within the PEIS-Home (either the robot's camera or any of the ceiling cameras). It is constituted by two components, the trainer and the recognizer. The trainer can handle multiple images per object and ultimately constructs a hierarchical database file with all the SIFT features extracted from the input images. In the structure of the trained database file one may find the features indexed by objects, models and bags of features (Fig. 2). For example, for an object that consists of several pictures from different viewpoints, a new object category is created that includes several models (essentially viewpoints), with each model containing bags of SIFT features from those viewpoints. With this technique, overlapping SIFT key-points were combined under the corresponding most general model, allowing for recognition of objects irrespective of viewpoint. Objects are recognized by processing the video feed in a frame by frame basis, extracting the SIFT features from the camera image, and then matching those features against the trained database using a κ-d tree matching algorithm.

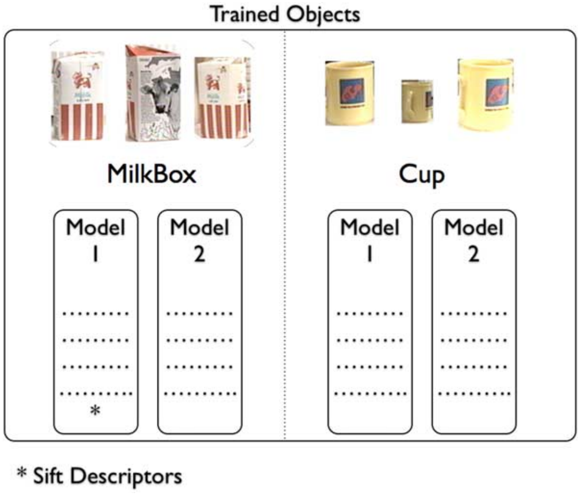

Fig. 2. Trained Database File.

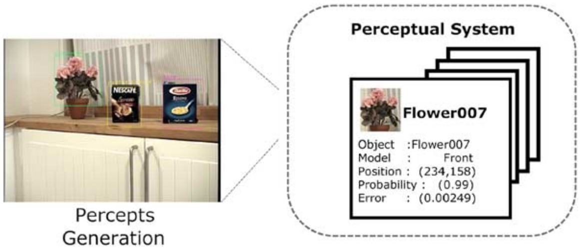

Fig. 3. Percepts formulation in the Perceptual Layer.

The object recognizer is able to recognize multiple objects in real-time, and in each time cycle, the perceptual signatures (explained in Section 2.2.1) from the recognized objects are published into the communication middleware. A perceptual signature, includes a latest copy (cropped area) of the corresponding object; its unique identifier and information about the scale, orientation, reference position, error, and probability of match. This is the driving force or stimulus to perceive the objects in the environment and the output is then available to the anchoring component for further information extraction. Figure 3 shows an example of the data contained in the perceptual system.

For localization and navigation used by the robot perceptual group (shown on the left of the perceptual layer), a number of components are networked. We use standard localization techniques and the detail on the implementation of these components are omitted. The other perceptual group which regards the smart home includes again the same visual processing algorithm explained above, and other sensors located in the environment. In the case studies (Section 3) we use a person tracking system which provides the position of the human in the environment. We also use small sensor devices in each room (Called Tmotes) which provide luminance, temperature and humidity readings.

2.2. *Physical symbol grounding*

The grounding of perceptual data presented in this work concern mainly household objects and in the presented system we use Perceptual Anchoring [9] to create and maintain in time the correspondence between symbols and percepts that refer to the same physical object. These modules consist of a grounding module, an anchor management, and memory management. The latter component is a novel addition which is used to facilitate the management of information to the knowledge base synchronizer that interfaces with the KR&R. More detail on these modules are given in the following Sections.

2.2.1. *Perceptual anchoring*

The task of anchoring is to create and maintain in time the correspondence between symbols and percepts that refer to the same physical object. This correspondence is reified in a data structure $\alpha(t)$, called an *anchor*. It is indexed by time as the perceptual system continuously generates new percepts; and the created links are dynamic, since the same symbol may be connected to new percepts every time a new observation of the corresponding object is acquired. So at each time instance t, $\alpha(t)$ contains a symbol identifying that object, a percept generated by the latest observation of the object, and a perceptual signature meant to provide the (best) estimate of the values of the observable properties of the object.

The main parts of anchoring are:

- A *symbol system* including a set $\mathcal{X} = \{x_1, x_2, \ldots\}$ of individual symbols (variables and constants), a set $\mathcal{P} = \{p_1, p_2, \ldots\}$ of predicate symbols, and an inference mechanism. In this work, the symbolic system is the Cyc implementation.
- A *perceptual system* including a set $\Pi = \{\pi_1, \pi_2, \ldots\}$ of possible percepts, a set $\Phi = \{\phi_1, \phi_2, \ldots\}$ of attributes, and perceptual routines. A percept is a structured collection of measurements assumed to originate from the same physical object; an attribute ϕ_i is a measurable property of percepts with values in the domain $D(\phi_i)$. Let $D(\Phi) = \bigcup_{\phi \in \Phi} D(\phi)$.
- A *predicate grounding relation* $g \subseteq \mathcal{P} \times \Phi \times D(\Phi)$, which embodies the correspondence between predicates and values of measurable attributes. The relation g maps a certain predicate to compatible attribute values.

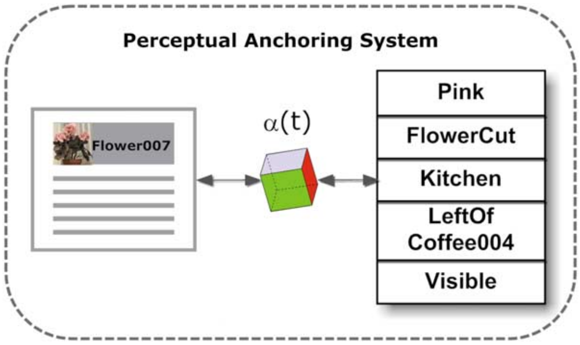

Fig. 4. Grounding of the perceptual signature to formulate an anchor in the perceptual anchoring layer.

The following definitions allow to characterize objects in terms of their (symbolic and perceptual) properties:

- A *symbolic description* σ is a set of predicates from $\mathcal{P}$.
- A *perceptual signature* $\gamma : \Phi \mapsto D(\Phi)$ is a partial mapping from attributes to attribute values.

The *predicate grounding relation* g is responsible for encoding the attribute values compatible with certain predicate symbols, while anchoring concerns encoding those relations with certain individuals' (object) symbols.

An example from the symbol system may indicate that in the individual symbols' set $\mathcal{X}$ there exist symbols such as $\{cup - 22, cup - 12, garbageCan - 2\ldots\}$ while the predicate symbols' set $\mathcal{P}$ may contain $\{small, big, red, black \ldots\}$. In the perceptual system, the set Π indicates percepts as $\{Image, HSV, Area, Position \ldots\}$, which correspond to values from the attributes' set Φ. Here it is important to mention that the *predicate grounding relation* g is responsible for encoding the attribute values compatible with certain predicate symbols, while anchoring concerns encoding those relations with certain individuals' (object) symbols.

2.2.2. *Grounding techniques*

To ground perceptual signatures from different sensing modalities an interface that allows to connect various grounding plugins together, is used where each plugin implements its own grounding relation. For the experiments of the current work, we address five different grounders regarding the input from the visual processing component that formulate a sufficient symbolic description of the objects located in the environment, from their perceptual signatures. Those are object category, color, location, spatial relations and visibility grounding. For color grounding we use a n-pixel random sampling algorithm applied to the segmented image of the recognized object, averaging the selected pixels, while the grounding relation maps the result between twelve classes of colors (1. Red, 2. Green, 3. Yellow, 4. Blue, 5. Black, 6. White, 7. Pink, 8. Cyan, 9. Gray, 10. Orange, 11. Brown, 12. Purple) that can be accurately differentiated by people with standard vision [6]. The location is derived by correlating the position of the robot or camera (using localization perceptual data) to the semantic map that is provided to the system. The object category is taken directly from the object recognition module as it is included in the unique identifier of the percept. An example of the anchoring system is illustrated in Fig. 4.

Spatial relations constitute an important symbolic description, because they allow humans to distinguish objects by their location with respect to other objects, possibly identical, and plays an important role when it comes to human-robot interaction. Two classes of binary spatial relations between a reference object and the located object are considered: the distance (topological) relations "at" and "near" and the directional (projective) relations "front of", "behind", "right" and "left". For reasons of simplicity we assume a deictic frame of reference with an egocentric origin coinciding with the robot platform. For the computation and evaluation of these spatial relations within the grounder plugin, we use the model presented in [7]. The computation of the spatial relations in the anchoring module have been validated previously in [8]. We arbitrarily assume that objects not observed at some point in time, do not change their location. Lastly, visibility is grounded by default as true during grounding, since all the percepts generated by recognized objects are visible at the time. This attribute changes in the functionality of the anchoring management described in the Section 2.3.

For the Tmotes of the smart home, fuzzy grounders for grounding temperature luminance and humidity are used. We use the information from the person tracking system together with the symbolic map provided to the location grounder.

2.3. *Management & perceptual memory*

The Anchoring Manager, is responsible for distributing the recently grounded anchor candidates to the rest of the system. This is achieved by extending and modifying the anchoring's functional part originally proposed in [9]. The anchoring process concerns bottom up and top down information acquisition through the following functionalities: find, acquire track and reacquire. *Find* is used for searching through the anchoring space. It searches both the symbolic descriptions and perceptual signatures. It can be triggered by the grounding in order to search if there is already an anchor created for the specific grounded object, or similarly it can be invoked by a user command trying to identify an anchor against its symbolic description (e.g. "red" & "cup"). If we receive a candidate anchor which does not match any existing anchor, *acquire* is used to initiate a new anchor for this candidate. Else, if the candidate matches an existing anchor, *track* assures that the perceptual signature pointed to by the existing anchor is the most recent and adequate perceptual representation of the object, by updating the perceptual signature with the one from the candidate anchor.

The novel functions is the *perceptual memory management*. It consists of two memory structures, one for storing anchors currently perceived by an active perceptual agent (e.g. robot) (working memory) and another one which stores previously perceived anchors (archive memory). If, for example, the robot is not perceiving an object anymore, the anchor of the corresponding object is marked as not visible and then archived (archive function). While if we happen to identify an object that we have previously anchored, the *reacquire* function brings up the corresponding anchor from the archive memory.

2.4. *Commonsense knowledge representation and reasoning*

A central contribution in this work is the KR&R and the necessary components to interface with the anchoring/grounding processes. The objective of this computational layer, is to transform the grounded symbolic information that exists in the Perceptual Anchoring layer, into hierarchically structured semantic information defining concepts at different levels of abstraction conforming with a shared vocabulary (Ontology). This information can be exchanged between agents, including humans without being dependent on an interpretation context while allowing reasoning. The role of the of the KR&R system is to maintain a world model which consists of the collection of the semantic information perceived by the robotic assistant, and the smart home. The commonsense KR&R system used in this work is Cyc, which contains roughly 3 million assertions about 250,000 concepts. Operations (such as assertions, retractions, modifications and queries) are stated using Cyc's formal language, CycL. Essentially CycL is an extension to second order predicate calculus, which is unambiguous and enable mechanical reasoning. The extensions also include features of higher order logics (e.g. quantification over predicates, functions, and sentences) that allow higher order assertions. Such a system, that rests on a large scale general purpose knowledge base, can potentially manage tasks that require world knowledge (or commonsense), the knowledge that every person assumes his neighbor also possess.

2.4.1. Knowledge base

The main challenge of integrating a large KR&R like Cyc is to be able to synchronize the information with the perceptual data coming from multiple sensors, which is inherently subject to incompleteness, glitches, and errors. The anchoring module provides a stable symbolic information despite fluctuating quantitative data through the grounding plugins. Nonetheless, instances of objects must be created in Cyc and as objects can be dynamic (e.g. in the case of changing properties) proper updating of information needs to be managed. Ultimately, to enable the synchronization between the KRR and anchoring layers three aspects are considered: defining anchoring in the KRR, handling of assertions, and handling of ambiguities.

It is important to define the context of anchoring as Cyc is not consistent globally but rather tries to be consistent locally by exploiting the use of different contexts which are expressed as *MicroTheories*. Through the Anchoring *MicroTheory* it is possible to connect concepts about objects that are currently present in the anchoring module, to the structured hierarchical knowledge in Cyc and concurrently inherit the common sense reasoning about this knowledge. For instance, if the location of "cup" stored in the anchor, is the Kitchen, then this object and the kitchen are instantiated into Cyc, inheriting all the properties related to the generalized concepts of the "kitchen" and the "cup", such as *Man Made Thing* or *Humanly Occupied Spatial Object*.

To enable the Anchoring *MicroTheory* (Mt), the following formulae were asserted:

1. Anchoring Mt is a general Cyc microtheory.
2. Everything true in "things known", independent of context, is also true in Anchoring Mt.
3. Everything true in Cyc's knowledge of general truths is also true in Anchoring Mt.
4. If through experience an intelligent agent perceives a SENTENCE, then that SENTENCE is a True Sentence.
5. If through experience an intelligent agent perceives a SENTENCE then that intelligent agent knows that SENTENCE.
6. If some localized spatial thing is located in some other localized spatial thing LOCAL and some other spatial thing REGION contains LOCAL, then that localized spatial thing is located in REGION.

Points 2 &3 are used in order to inherit all that Cycknows about the world. Point 4 (Second Order Assertion) is used to make the agent's perception true inside the AnchoringMt. In addition, it is also necessary to assert the concept of an agent. This has been done by initially creating the agent given a specific name (e.g. Self) and specifying a number of is-A assertions about the agent such as:

– Self is a localized spatial thing.
– Self is a tangible agent.
– Self is a perceptual agent.

Furthermore, the concept of an anchor and corresponding statements were defined (e.g. an anchor is a data structure, it indicates some instance of an object, which has some properties like color, location, spatial relations, ...).

2.4.2. Knowledge synchronization

To synchronize the knowledge in Cyc with the perceptual data, instances of the anchored objects present in both working and archive memory are asserted. The knowledge synchronization's role keep the Cyc symbolic system coherent with the symbolic descriptions of each anchor. This is done by translating the symbolic description of each anchor into a set of local formulae that state the agent's perception about the object. Figure 5, exemplifies how assertions are made. When a particular predicate of the symbolic description changes (e.g. when the location of an object is in

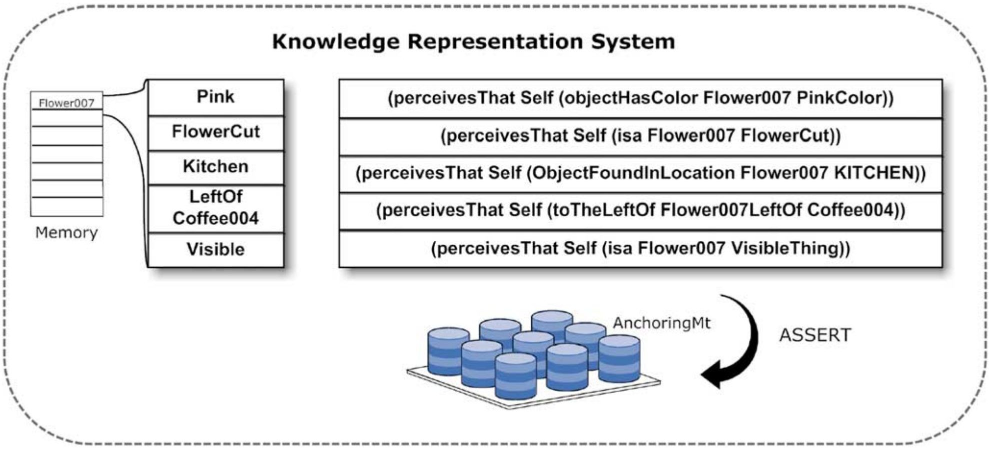

Fig. 5. Transformation of the anchor's symbolic part, to logic formulae assertions about the agent's perception of the corresponding anchor (object).

another room), first an unassertion of the predicate is made, and a new assertion containing the updated information is performed.

Ambiguities may arise during the synchronization process and are resolved using the *denotation tool* interactively with the user. The denotation tool is an interactive step during the grounding process where the user is asked to disambiguate the symbol that is grounded between all the concepts in the KB that denote this symbol. This is exemplified here. As an example, when the anchoring of a "flower" triggers a KB assertion about this flower, an ambiguity will arise as Cyc contains multiple concepts denoting the term "Flower"; the possible options are: `CutFlower, Flower-BotanicalPart, FlowerDevelopmentEvent, FloweryPlant`. The user is asked to select between all those Concepts, which are sorted by the probability of relevance.

Another aspect to synchronize is the information about the visibility of the object. This is important as the physically embedded agent still needs to maintain knowledge about an object, despite that the corresponding perceptual information is not currently available to the agent. For example the robot still needs to maintain the knowledge that "a cup is located in the kitchen", despite that the "cup" is outside its current field of view. This is achieved by 2^{nd} order rule assertions as stated in Section 2.4.1 and an example is shown in the experiments in Table 2.

2.5. NL communication

To combine Cyc's natural language capabilities and enhance the communication with the user we allow

Table 1
NL Template Examples

```
-What do you see ?
query:(sees AGENT ?OBJECT)
-Where is the red cup ?
query:(objectFoundInLocation
        (and (isa ?CUP Cup)
        (objectHasColor ?CUP RedColor)
       ?WHERE))
-Do you know that the green flower is located in the kitchen?
query:(knows AGENT
  (and (isa ?FLOWER Flower)
  (objectHasColor ?FLOWER GreenColor)
  (objectFoundInLocation ?FLOWER KITCHEN)))
```

Cyc to translate the results of queries or inferences, into natural language. In addition a console based graphical user interface is created allowing the user to type in questions or complementary assertions in natural language which are parsed by a template based NL parser, and in turn accesses Cyc to perform the required operations. All the occurrences of the the word 'you' are reformulated to the agent's name. The NL parser is able to capture 15 categories of *what* questions, 19 categories of *who* questions, 7 categories of *is* questions and 18 other categories. In addition there are 11 other possible commands which include assertion types (set, *is*-A, generalization, assert, make), informational types (get, complete, explain, translate, guess-denote), retraction types (retract, delete). Some examples of how this translation are given in Table 1.

While from a technical point of view, communication occurs between the human user and the PEIS-

Fig. 6. From left to right. Sample training images of objects. The robot's perceptual space with the visually recognized objects. Similarly, ceiling camera's perceptual space. The PEIS-home Symbolic Map.

Ecology, it is presented to the user as an interaction between the user and the robot, as this interaction has shown to facilitate a better HRI.

3. Case studies

The implemented architecture is built upon a communication framework that allows the dynamic connection between the different components described in the previous sections. A component is any computerized system interacting with the environment through sensors and/or actuators and including some degree of "intelligence". Each component is called a PEIS or physically embedded intelligent system and can be as simple as a smart toaster and as complex as a humanoid robot. A PEIS-Ecology is the collection of all those components. In our realization of a PEIS-Ecology, the PEIS relies on a distributed middleware to communicate and cooperate. This PEIS-middleware implements a distributed tuple-space on a P2P network: PEIS exchange information by publishing tuples and subscribing to tuples, which are transparently distributed by the middleware. Each PEIS also provides a set of standard tuples, e.g., to announce its physical appearance or the functionalities that it can provide.

As part of the testbed, we use a physical facility, called the PEIS-home, which looks like a typical bachelor apartment of about $25m^2$. It consists of a living room, a bedroom and a small kitchen as shown in the Fig. 6 (right). The PEIS-Home is equipped with communication and computation infrastructure along with a number of sensors like camera's and localization components. In our ecology there is an interactive robotic system based on an ActiveMedia PeopleBot platform for indoor research. In addition to the usual sensors, the robot is equipped with a SICK LMS200 laser range finder and with a Sony PTZ color camera.

3.1. Perceptual information & properties

We initially presented the robot with 15 objects while capturing 2 to 5 training images per object from different viewpoints. Some training images samples are shown in Fig. 6 (left). It is important to mention that the robot recognizes instances of objects and not object categories. We placed those objects around the PEIS-home in a random manner, either close or far, covering a great amount of different combinations of spatial relations. We then allowed the robot to navigate around the environment so as to recognize and anchor those objects. We then allowed the human to communicate with the system, by typing into the console based graphical user interface, questions (related to robot's perceptions, or commonsense), or communication commands in natural language according to the NL part described in Section 2.5. An excerpt of the dialogue that reflects perceptual information from the mobile robot and the environment is presented in Table 2. At the point where this dialogue took place, the robot was located in the kitchen looking at a milk box, a washing liquid and some tea, as indicated in Fig. 6 (2nd picture) which is the output of the object recognition component. We are also able to use the visual sensors which are located in the ceiling and can observe things that the robot would not be able to observe because they are too large[1] or occluded. Another instance of the object recognizer was running on the large objects from the PEIS-home (e.g. television set, couch, table, sink, etc.). Through the knowledge base assertions which pertained a larger collection of objects of the environment, the robot is able to compensate for its limited ability of perception.

[1]Due to PEIS-home space restrictions the robot is not able to position it self far enough in order to observe the whole couch or the whole television set.

Table 2

Excerpt from the dialogue between the human and the Anchoring Robot

User >	What do you see?
Self >	I perceive a Milk Box, a Washing Liquid and a Tea.
User >	Where are you?
Self >	I am located in the Kitchen.
User >	What color is the Washing Liquid?
Self >	Green.
User >	Where is the Television Set?
Self >	The Television Set is located in the LivingRoom.
User >	What Color is Television Set?
Self >	Black.
User >	What is the Temperature in the Kitchen?
Self >	The Temperature in the Kitchen is 21C.

Table 3

Qualitative Spatial Relations dialogue

User >	Where is the Pink Flower?
Self >	The Pink Flower is located in the Kitchen, to the left of Coffee, to the Left of Soup, (...) and to the right of Me.
User >	Where is the medicine?
Self >	The Anti Inflammation Gel in front of Me, or the Painkiller to the right of Me?

3.2. *Qualitative spatial relations*

Humans heavily use qualitative descriptions about the spatial relations of the objects they communicate. The perceptual grounded knowledge in the KB contains also the spatial relations of each object. In this scenario we demonstrate that spatial relations like “nFrontOf”, “Behind”, “ToTheLeft” are correctly generated by the system, and are used for obtaining spatial descriptions of objects that have been previously anchored or even disambiguate between instances of concepts that fall into the same generalization like in the example in the Table 3. Notice also in the last query of the dialogue that the ontology within the KB is exploited to provide for a more robust dialogue and the ambiguity is possible to resolve by providing more information about the spatial relations of the objects.

Table 4

Commonsense Queries

User >	What Things are capable of flying?
Self >	Bird, Airoplane, (...) are capable of flying.
User >	Is the Flower capable of flying?
Self >	I cannot prove that.
User >	Do you know that the yellow cup is located in the kitchen?
Self >	Yes.
User >	Do you perceive that the yellow cup is in the kitchen?
Self >	No.
User >	What is a Flower?
Self >	Flower007 is a Flower. Flower is an Organism Part, a Finite Spatial Thing, a Biological Living Object, a Partially Tangible Thing, a Three Dimensional Thing, (...).
User >	Is the Flower in the Kitchen, a Bird?
Self >	I cannot prove that.

3.3. *Commonsense reasoning*

The advantage of using a KB that contains commonsense information is that we can exploit this information to be able to infer things that were not directly asserted perceptually. This information can be useful are when queries about functions and properties of objects are made. Due to the expressiveness of Cyc we could manage to differentiate perceptual and epistemic knowledge, but at the same time keeping both coherent. In the example in Table 4, we can see that the system can support both queries of “knowing that ...”, and “perceiving that ...”. This virtually creates the effect of memory, where we know about things we have seen in the past, but we currently do not perceive any more. This function is supported by the memory structures in the Perceptual Anchoring level, where it is explicitly stated what is perceived or not, at any given point in time. What is not perceived any more, the perceptual information is not discarded but transferred to knowledge repository (e.g. as explained in Section 2.4.1). In this way, this information to be accessible in the future. When objects are not perceived, the assumption is that their properties do not change, however, new perceptual information can update an objects properties.

4. Conclusion & future work

In this work we have investigated the integration of Cyc with a perceptual system consisting of networked sensors. The common sense functionality in Cyc was exploited to reason about objects and their properties. The use of Cyc offered advantages, as there was less time spent on ontology concerns as well as entering or creating new knowledge in the KB. A further advantage is the exploitation of commonsense knowledge in order to infer things that could not be inferred by having the symbolic descriptions on their own. Focus was concentrated on the perceptual anchoring and the necessary processes to synchronize the symbolic system with the perceptual data. The dialogue has been used to validate the system. To our knowledge this is the first integration attempt of a substantial deterministic logical database with a perceptual system with real world sensed data. It is the combination of all three, robust perceptual system, anchoring and a substantial KB that enables us to ground NL oriented queries in observations through the procedure that is explained in this paper, in a dynamically changing environment.

A number of issues will be considered for future validation. There are some practical considerations regarding the improvement of the current architecture. We find that the object recognition component can be greatly improved by a GPU assisted version of the SIFT algorithm [5] to achieve close to realtime performance. We plan to replace the model for computing the spatial relations with the one found in [25], so as to approximate the global positions of the recognized objects and then be able to compute a larger set of spatial relations, including objects not seen, at some point in time. Most importantly we anticipate an improvement of the natural language understanding mechanism by a more sophisticated method which uses the full spectrum of Cyc's natural language knowledge and capabilities, so as to be more flexible and robust when translating from natural language to logic formulae. It would significantly help also in the case where we need to teach about new concepts in the Knowledge Base using spoken language. Finally with respect to the evaluation, we are interested to investigate the behavior of the system when scaling some factors, for instance the number of objects, or the grounding plugins.

References

[1] C. Galindo, J. Fernandez-Madrigal, J. Gonzalez, and A. Saffiotti, "Robot task planning using semantic maps", *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 955–966, 2008.

[2] O. Mozos, P. Jensfelt, H. Zender, G.-J.M. Kruijff, and W. Burgard, "An integrated system for conceptual spatial representations of indoor environments for mobile robots", in *Proceedings of the IROS 2007 Workshop: From Sensors to Human Spatial Concepts (FS2HSC)*, San Diego, CA, USA, November 2007, pp. 25–32.

[3] G.-J.M. Kruijff, P. Lison, T. Benjamin, H. Jacobsson, and N. Hawes, "Incremental, multi-level processing for comprehending situated dialogue in human-robot interaction", in *Symposium on Language and Robots*, Aveiro, Portugal, 2007.

[4] H. Holzapfel, D. Neubig, and A. Waibel, "A dialogue approach to learning object descriptions and semantic categories", *Robot. Auton. Syst.*, vol. 56, no. 11, pp. 1004–1013, 2008.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[6] B. Berlin and P. Kay, *Basic Color Terms: Their Universality and Evolution*. Center for the Study of Language and Inf., March 1999.

[7] K. Gapp, "An empirically validated model for computing spatial relations", in *KI '95: Proceedings of the 19th Annual German Conference on Artificial Intelligence*. London, UK: Springer-Verlag, 1995, pp. 245–256.

[8] A. Loutfi, S. Coradeschi, M. Daoutis, and J. Melchert, "Using knowledge representation for perceptual anchoring in a robotic system", *"Int. Journal on Artificial Intelligence Tools"*, vol. 17, pp. 925–944, 2008.

[9] A. Loutfi, S. Coradeschi, and A. Saffiotti, "Maintaining coherent perceptual information using anchoring", in *The Nineteenth International Joint Conference on Artificial Intelligence (IJCAI '05)*, Edinburgh, Scotland, 2005, pp. 1477–1482.

[12] Dejan Pangercic, Rok Tavcar, Moritz Tenorth, and Michael Beetz. Visual scene detection and interpretation using encyclopedic knowledge and formal description logic. In *Proceedings of the International Conference on Advanced Robotics (ICAR).*, 2009.

[11] Mykel J. Kochenderfer and Rakesh Gupta. Common sense data acquisition for indoor mobile robots. In *Nineteenth National Conference on Artificial Intelligence (AAAI-04*, pages 605–610. AAAI Press / The MIT Press, 2003.

[12] Dejan Pangercic, Rok Tavcar, Moritz Tenorth, and Michael Beetz. Visual scene detection and interpretation using encyclopedic knowledge and formal description logic. In *Proceedings of the International Conference on Advanced Robotics (ICAR)*, 2009.

[13] Paulo E. Santos. Formalising the common sense of a mobile robot. Technical report, 1998.

[14] Murray Shanahan. Robotics and the common sense informatic situation, 1996.

[15] Moritz Tenorth and Michael Beetz. Towards practical and grounded knowledge representation systems for autonomous household robots. In *Proceedings of the 1st International Workshop on Cognition for Technical Systems, München, Germany, 6-8 October*, 2008.

[16] Michael Witbrock, David Baxter, Jon Curtis, Dave Schneider, Robert Kahlert, Pierluigi Miraglia, Peter Wagner, Kathy Panton, Gavin Matthews, and Amanda Vizedom. An interactive dialogue system for knowledge acquisition in cyc. In *Proceedings of the Workshop on Mixed-Initiative Intelligent Systems. IJCAI*, pages 138–145, 2003.

[17] Hendrik Zender and Geert-Jan M. Kruijff. Multi-layered conceptual spatial mapping for autonomous mobile robots. In Holger Schultheis, Thomas Barkowsky, Benjamin Kuipers, and Bernhard Hommel, editors, *Control Mechanisms for Spatial Knowledge Processing in Cognitive / Intelligent Systems*, volume Technical Report SS-07-01 of *Papers from the AAAI Spring Symposium*, pages 62–66, Menlo Park, CA, USA, March 2007. AAAI Press.

[18] Hendrik Zender and Geert-Jan M. Kruijff. Towards generating referring expressions in a mobile robot scenario. In *Language and Robots: Proceedings of the Symposium*, pages 101–106, Aveiro, Portugal, December 2007.

[19] A. Bonarini, M. Matteucci, and M. Restelli, "Concepts for anchoring in robotics", in *AI*IA 2001: Advances in Artificial Intelligence*, ser. Lecture Notes in Computer Science, F. Esposito, Ed. Berlin, D: Springer Verlag, 2001, vol. 2175, pp. 327–332.

[20] G.-J.M. Kruijff and M. Brenner, "Modelling spatio-temporal comprehension in situated human-robot dialogue as reasoning about intentions and plans", in *Symposium on Intentions in Intelligent Systems, AAAI Spring Symposium Series 2007, Stanford University, Palo Alto, CA, USA*, 2007.

[21] C. Matuszek, J. Cabral, M. Witbrock, and J. DeOliveira, "An introduction to the syntax and content of cyc", in *AAAI Spring Symposium*, 2006.

[22] A. Saffiotti and M. Broxvall, "PEIS ecologies: Ambient intelligence meets autonomous robotics", in *Proc of the Int Conf on Smart Objects and Ambient Intelligence (sOc-EUSAI)*, Grenoble, France, 2005, pp. 275–280.

[23] Harnad, S. (1990). The symbol grounding problem. *Physica D.*, *42*, 335–346.

[24] Kevin LeBlanc and Alessandro Saffiotti. Cooperative anchoring in heterogeneous multi-robot systems. In *ICRA*, pages 3308–3314. IEEE, 2008.

[25] Marjorie Skubic, Dennis Perzanowski, Alan C. Schultz, and William Adams. Using spatial language in a human-robot dialog. In *ICRA*, pages 4143–4148. IEEE, 2002.