Knowledge Representation for Anchoring Symbolic Concepts to Perceptual Data

Marios Daoutis, Amy Loutfi, Silvia Coradeschi

Abstract Perceptual anchoring is the process of creating and maintaining a connection between the sensor data corresponding to a physical object and its symbolic description. It is a subset of the general symbol grounding problem and has been investigated over the past years. In this chapter we present a method for grounding sensor data of physical objects to the corresponding semantic descriptions, in the context of cognitive robots. Specifically we investigate the challenge of establishing the connection between percepts and concepts referring to objects, their relations and properties. We examine how knowledge representation can be used together with an anchoring framework, so as to complement the meaning of percepts and support better linguistic interaction with the use of concepts. This implies that robots need to represent both their perceptual and semantic knowledge, often expressed in different abstraction levels while originating from different modalities. We focus on the integration of anchoring with a large scale knowledge base system and with perceptual routines. This integration is applied in a number of studies, which span from high-level to commonsense reasoning. A specially interesting application of anchoring is in the context of smart home and we consider it as an example of the applicability of the anchoring framework.

1 Introduction

Artificial systems combining perception and action with high-level conceptual knowledge are essential for Human-robot Interaction (HRI) applications in real world and dynamic environments. In such applications humans are considered an integral part and robots become a true partner cooperating towards a common goal.
An important requirement for systems interacting with humans is a common understanding of the world in which they both operate. This requirement includes sharing and communicating knowledge referring to the physical entities from the environment, such as objects, rooms and locations. This knowledge is partly common-sense knowledge and partly specific perceptual knowledge regarding a current situation. Moreover it is typically symbolic in nature as it is expressed in natural language.

However, intelligent systems need a great quantity of knowledge for a specific task which is often hard to provide a-priori, even in simple problems. In addition general knowledge needs to be complemented with specific knowledge regarding objects in the environment (i.e. how the object has been used, where it has been placed and how their location changes over time).

As the artificial systems progress steadily from sensor based reactivity, to cognitive computational models that enable them to perceive, learn, reason and interact, we need to deal with issues such as: a) how sensory and perceptual information is acquired and grounded to symbols; b) how the meaning of the symbols is represented; and c) how to express this meaning within a vast amount of formally defined commonsense knowledge.

We explore the problem of associating the perception of physical objects with the corresponding conceptual knowledge and common-sense information, in the context of cognitive robots. We approach this problem via a mechanism that establishes the link between perception and conceptual knowledge, which is studied in the field of Anchoring [17], where anchoring is the process to establish and maintain in time the correspondence between symbolic knowledge and sensory data.

1.1 Anchoring in Grounded Interaction

Our focus is on the problem of anchoring in a cognitive robotics scenario. To get a better understanding of how anchoring is involved even in very simple cases, consider the example shown in Fig. 1. A cognitive robot observes a novel object and speculates according to its prior perceptual knowledge that the object in discussion might actually be a “mug”. We can easily identify already, that the robot needs to be able to sense and perceive a complete and coherent representation of the sensed object, while binding its features together, despite the eventual uncertainty. Then it attempts to associate the novel object with its past perceptual experiences in an effort towards the identification of the object. It appears quite trivial when the human asserts that he owns this “coffee-cup”, however the linguistic assertion presupposes the presence of related knowledge as well as the corresponding semantics of this knowledge. Below we see more in detail the underlying principles of this motivational scenario.
1.2 Chapter Overview

The research behind this chapter is focused on anchoring and specifically its application to cognitive systems. Although the basic elements of an anchoring system were formally introduced in various articles [13, 14, 17, 37], several aspects have been left out of focus. One such aspect concerns the design and implementation of a cognitive perceptual system that is based on an anchoring framework for its computational model.

More specifically in this chapter we elaborate on a) how the different perceptual modalities and grounding mechanisms can be modelled in such a system; b) how its perceptual-semantic knowledge is defined and represented; and c) how a-priori knowledge can be used as cognitive bias. In addition we intend to present the capabilities of such a system while identifying the benefits of using anchoring as the main computational model.

In this chapter, we leverage from the existing research on anchoring and we provide a framework for the concept-percept correspondence, which accommodates and seamlessly integrates general knowledge with perceptual information acquired by the system about the relevant objects that are present in the environment. Even though the framework is primarily intended as a solution to the Symbol Grounding Problem [25] in reality we consider how a cognitive agent (i.e. a robot) may acquire novel concepts and reason about these concepts perceptually and semantically.

We first present the related work (§2) and the anchoring framework (§3), while then we consider how anchoring can be used together with known knowledge representation techniques, in the context of spatial (§5) and commonsense (§6) reasoning. We then consider a cooperative aspect where several agents interact with the human during the anchoring process (§7). Finally we consider how anchoring can be per-

---

1 Symbol Grounding denotes the grounding of symbols to their sensorimotor representations.
formed when using information from the web as an additional source of knowledge (§3). The aim of these parts is to provide a coherent overview of our current work in anchoring and its progression.

2 Perceptual Anchoring in Cognitive Systems

The anchoring problem was defined in [14] and we have considered a variety of aspects related to this problem including an extension of the anchoring framework encompassing actions [15] and planning [16]; the validation of anchoring in several domains [17]; anchoring under uncertainty [11]; the study of anchoring in a new domain of electronic olfaction [35]; the development of a first system integrating planning and anchoring in order to detect and recover from ambiguities by performing perceptual actions [28] and anchoring of maps [23]. In the literature the study of anchoring led different approaches to address the problem from diverse perspectives. Here we present the most notable clusters of work on anchoring, related also to the aspects of this chapter.

2.1 Knowledge Based approaches to Anchoring

There exist many approaches in the context of physical symbol grounding which consider the use of KR&R techniques or knowledge bases, in order to enable logical inference. Work by Bonarini et al. studies a model to represent the knowledge of an agent, showing that the anchoring problem can be successfully dealt with well known AI techniques. They present a model which supports the instantiation of concepts, affected by uncertainty and heterogeneity from the perceptual system in a multi-agent context [2, 3].

One of the earliest clusters of work in the context of knowledge based approaches, is led by Chella et al. where they present a knowledge based anchoring framework, for building high-level conceptual representations from visual input, based on the notion of Gardenfors’ conceptual spaces [24]. Their approach allows to define conceptual semantics for the symbolic representations of the vision system, where the symbols can be grounded to sensor data. A collaboration with Coradeschi and Saffiotti [6], explicitly investigates how to formalise a computational model for perceptual anchoring that is unified with Gardenfors’ conceptual spaces theory, in the context of bridging the gap between the symbolic and sub-symbolic components.

Another interesting knowledge-based approach for cognitive robots, is the one from Shapiro and Ismail called the GLAIR architecture (Grounded Layered Architecture with Integrated Reasoning) which consists of three levels. The knowledge level (KL) is the one in which conscious reasoning takes place [46]. The KL is implemented by the SNePS and SNeRE (SNePS Relational Engine) logic & knowledge representation and reasoning system [30, 47], both based on Common Lisp.
They evaluate their approach using the robot Cassie, which anchors the abstract symbolic terms that denote the agent’s mental entities in the domains of: a) perceivable entities and properties; b) actions; c) time; and d) language in the lower-level architectures used by the embodied agent to operate in the real world.

Mozos et al. present an integrated approach for creating conceptual representations for spatial and functional properties of typical indoor environments using mobile robots [42]. Their multi-layered model represents maps at different levels of abstraction, using laser and vision sensors for place and object recognition. Their system is endowed with an OWL-based commonsense ontology of an indoor environment, that describes taxonomies (is-A relations) of room types and typical objects found therein, through has-A relations. Zender et al. present another integrated instance of the system with functionalities such as perception of the world, natural language, learning and reasoning, exploiting inter-disciple state-of-the-art components into a mobile robot system (CoSy Explorer) [52]. Their work is highly focused on cross-modal integration, ontology-based mediation and multiple levels of abstraction of perception.

2.2 Anchoring with Commonsense Information

Of the most notable related approaches is the one by Tenorth et al., which concerned household robots that use commonsense knowledge to accomplish everyday tasks. They present their integrated system which focuses on the generation of complex plans. They propose to transform task descriptions from web sites such as ehow.com, into executable robot plans by using methods for converting the instructions from natural language into a formal logic-based representation, while then resolving the word senses using the WordNet lexical database and the Cyc ontology [51].

Another work which considers a commonsense ontology is presented by Lemaigre et al., in a knowledge processing framework for robotics called OpenRobots Ontology kernel (ORO) that allows to turn previously acquired symbols into concepts, linked to each other, thus enabling reasoning [33]. Knowledge in ORO is represented as a first-order logic formalism, in RDF triples (e.g. <robot isIn kitchen>) in OWL-DL2 and the knowledge base is implemented using the Jena semantic web framework in conjunction with the Pellet reasoner. The OpenRobots Common Sense ontology is closely aligned with the open-source OpenCyc upper ontology defining classes (56 are currently defined) and predicates (60 are currently defined) focused on concepts useful for interaction with humans.

\[2\] Cognitive Systems for Cognitive Assistants - CoSy, \url{http://www.cognitivesystems.org}

\[3\] Jena Semantic Web Framework, \url{http://jena.apache.org/}

\[4\] Pellet: OWL 2 Reasoner for Java, \url{http://clarkparsia.com/pellet/}

\[5\] OpenCyc KB, \url{http://www.opencyc.org}
2.3 Cooperative Anchoring

The most related cluster of work in cooperative anchoring, mainly comes from [LeBlanc and Saffiotti] where the authors propose an anchoring framework for both single-robot and cooperative anchoring. They primarily consider the problem of fusing pieces of information from a distributed system, to create the global notion of an anchor in a shared multi-dimensional domain, called the anchoring space. The framework represents information using a conceptual spaces [24] approach, allowing various types of object descriptions to be associated with uncertain and heterogeneous perceptual information. Their implementation uses fuzzy sets to represent, compare and combine information, while includes a cooperative object localisation method that takes into account uncertainty, in both observations and self-localisation.

Another instance of work is presented by Bonarini et al. [4] where the authors extend their solution of single-robot symbol grounding into a multi-agent approach, thus combining the information from different agents in a global representation at the conceptual level using a fusion model based on clustering techniques. Also Mastrogiovanni et al. present a distributed knowledge representation and data fusion system for an ambient intelligent environment, which consists of several cognitive agents with different capabilities. The architecture itself is based on the idea of an ecosystem of interacting artificial entities, where the framework for collaborating agents allows them perform an intelligent multi-sensor data fusion. Despite its simplicity, it is able to manage heterogeneous information at different levels, thus “closing the loop” between sensors and actuators.

2.4 Anchoring for Human-Robot Interaction

Chella et al. present a system for advanced verbal interactions between humans and artificial agents with the aim to learn a simple language in which words and their meaning are grounded in the sensory-motor experiences of the agent. The system learns grounded language models from examples with minimal user intervention and without feedback. Then, it has been used to understand and subsequently to generate appropriate natural language descriptions of real objects, engaging in verbal interactions with a human partner [8].

Another example of an implemented dialogue system is explored in [29, 38], in the context of situated dialogue processing. This approach couples incremental processing with the notion of bidirectional connectivity, inspired by how humans process visually situated language in both top-down and bottom-up fashion. Information about the object state as well as a history of the object state is used to describe changes in a scene.
2.5 The Web as Information Resource

Recent works begin to consider the web as an alternative source for acquiring information for use in robotics. Notable examples include the semantic robot vision challenge, an effort towards robots searching the environment for objects that had been learned through images retrieved from the web. There are also initial works using knowledge from the semantic web, or semantic web tools and technologies to provide knowledge to robots that operate in cooperation with the human [26]. However, even if the development of sophisticated commonsense knowledge bases is a quite active research topic with interesting results (E.g. Open Mind Common Sense[49], WordNet[22], YAGO[49], ConceptNet or CYC[34]), little attention is given in general on how to integrate those with perception in real world applications. In addition, a great effort is dedicated to map the CYC concepts, with Wikipedia concepts[41], with WordNet[44] and most importantly with DBpedia[44] the semantic translation of Wikipedia.

2.6 Summary of Related Approaches

Comparing the related clusters of work, we primarily see that even though there is much heterogeneity in the domains that study or relate to anchoring, only a few approaches follow an integrative stance when grounding symbols to sensor data in cognitive systems. Work in the cooperative anchoring domain seems to lean toward low-level sensing and grounding, thus emphasising on uncertainty and the multi-modal aspects. However, the approaches appear to miss the high-level symbolic aspects related to knowledge representation and reasoning [31] with the exception of Mastrogiovanni et al. [40] and their work on the distributed data fusion approach [40]. This approach explicitly considers symbolic knowledge representation and reasoning. We observe that integrative approaches which consider anchoring, knowledge grounding and symbolic reasoning, may rely on diverse representations of knowledge which vary according to the application. Gardenfors’ conceptual spaces representations are generally favoured in the context of vision and anchoring [6, 7, 9, 10, 31].

On the semantic side, mainly description logics are routinely used as the logical formalism behind the representation of knowledge [33, 50, 53]. Even though some knowledge-based models intended for cognitive robots introduce the concept of commonsense information in specific contexts, such as everyday tasks [51], important aspects like the multi-modal integration with cross-modal representations; memory and grounding, are not examined in detail. Exception is the ORO knowledge management platform [33] which explicitly deals with grounding linguistic interaction. In the context of linguistic interaction, anchoring approaches study pri-

---

6 The DBpedia project aims to extract structured information from Wikipedia, making it accessible on the Web. Currently, the knowledge base is considered as a central interlinking hub for the emerging Web of data [1]
marily the generation of referring expressions and spatial relations. Finally approaches tend to focus on the logical representations instead of the practical perceptual aspects behind the grounding and anchoring. However with the advancement of semantic web technologies and online commonsense knowledge repositories, we can only expect in the future, knowledge-enabled robots which eventually communicate and use grounded representations.

3 A Framework for Multi-Modal Semantic Anchoring

In our work we adopt and extend the basic anchoring model presented in [14]. In this section we describe the details and functionalities of the augmented anchoring framework, mainly developed in [18]-[21], [36].

The main aim of anchoring is to systematically create and maintain in time the correspondences between symbolic concepts and physical perceptual data that refer to objects from the physical environment. The anchoring system is composed of one or more perceptual systems, a semantic system and a set of semantic grounding relations which ground the percepts from the perceptual system(s) to the semantic system.

- **Perceptual System** Is mainly considered as a group of sensors and feature extraction algorithms which continuously produce percepts and attributes (measurable properties of percepts) assumed to originate from the same physical object.

- **Semantic Grounding Relation** Embodies the correspondence between the unary predicates in the semantic system and the attributes in the perceptual system.

- **Semantic System** Is the collection of all abstract, hierarchically structured, symbolic knowledge which is expressed in some form of formal logic and contains the set of concepts, instances as well as sets of relations and rules between the concepts.

The associations between symbols and percepts are reified via structures called anchors. Each anchor contains ontological fragments from the semantic system, percepts and estimates of objects’ properties. Anchors are time indexed, since their contents can change over time and are manipulated with five basic functionalities: a) find; b) (re)-acquire; c) track; d) query; and e) augment. Anchors contain both perceptual and semantic representations therefore we can describe objects in terms of their symbolic and perceptual properties. In addition the functionalities are able to handle both representations in the same way regarding matching.

The relations between concepts and perceptual instances are defined using the measure of similarity between anchors. In order to model this measure we use near sets. Near sets is a framework introduced by Peters and Ramanna, for solving problems based on perception, by defining the resemblance between sets. The Near sets theory is a generalization of rough sets and emerged from the idea...
that two or more rough sets can share objects with matching descriptions if they both contain objects belonging to the same class, and therefore are considered near. Nearness is defined as closely corresponding to, or resembling an *original* and near sets model the measures of nearness of sets based on similarities between classes contained in coverings of disjoint sets. In this context the term similarity means, resemblance between two concepts or between a concept and an instance where almost equal patterns are found in compared items. The computational framework we present below is a combination of instances of our previous work from [19] and [20].

### 3.1 Computational Model

Here we give a formal definition of the elements mentioned above, that allow us to characterise an object \( x \) in terms of its semantic and perceptual properties. We consider an *environment* \( E \) in which there exist a set \( O \) of \( m \) physical objects, where \( x_i \in O \) for all \( i = 1, \ldots, m \). The set of objects \( O \) can change over time, since objects can be inserted or removed from the environment.

There are also \( n > 0 \) perceptual agents, denoted as \( ag = ag_1, \ldots, ag_n \). If \( n = 1 \), then the problem is reduced to single-robot anchoring while for \( n > 1 \) we deal with the cooperative anchoring problem. Each perceptual agent \( ag \) has a number of sensors \( (S_{ag}) \) which produce perceptual data either in raw sensor measurements, or via feature extraction, feature vectors in a feature space.

Each object must have some quantifiable *features* (observable properties) that can be measured. So there exists a real valued feature function \( f : O \rightarrow \mathbb{R} \) which represents some feature of an object \( x_i \). All feature functions belong to the set \( F \) of the available feature functions of the perceptual system.

\( B \) is the set of all perceptual modalities of the perceptual system and can be modelled as \( B = (\beta_1, \beta_2, \ldots, \beta_k) \subseteq F \) where \( k \) is the number of modalities. For instance \( \beta_1 \) can be the modality of vision which concerns some feature functions \( \beta_1 = (f_1, f_2, f_3) \) which may represent *colour*, *texture* and *shape* features respec-
tively. Except the visual modalities, some other examples may include the olfactory, spatial or temporal, which can be overlapping sets of feature functions from $\mathcal{F}$.

### 3.2 Perceptual System

The perceptual system $\Pi$ is $\langle O, \mathcal{B}, \mathcal{F} \rangle$ and it consists of the non-empty finite sets of objects $O$, perceptual modalities $\mathcal{B}$ and available feature functions $\mathcal{F}$. Each perceptual system continuously produces real valued, structured collections of measurements (feature vectors) assumed to originate from the same object $x$. These collections of measurements are called percepts and are denoted as $\pi_\beta(x)$. The percept $\pi_\beta(x)$ of modality $\beta$ is a partial observation of an object $x \in O$, so that $\pi_\beta(x) = (\pi_{f_1}^\beta(x), \pi_{f_2}^\beta(x), \ldots, \pi_{f_i}^\beta(x))$ with $i$ being the number of feature functions $\in \beta$ and each $\pi_{f_i}^\beta(x) : f_i^\beta(x) \rightarrow \mathbb{R}$. The collection of percepts from all the modalities of the perceptual system with respect to an object $x$ is called the perceptual signature.

**Definition 1 (Perceptual signature).** The perceptual signature $\pi$ of an object $x \in O$ is defined as: $\pi_B(x) = \{\pi_{\beta_1}(x), \pi_{\beta_2}(x), \ldots, \pi_{\beta_k}(x)\}$.

Each $\pi_{\beta_i}(x)$ corresponds to a percept, or combined feature vector regarding a modality $\beta_i \in \mathcal{B}$. The perceptual distance between two perceptual signatures ($d_{\pi}$) of two objects $x$ and $y$, with $x$ and $y \in O$, is defined as:

$$
d_{\pi}(y-x) = \|\pi_B(y) - \pi_B(x)\| = \|\pi_{\beta_1}(y) - \pi_{\beta_1}(x)\|_2 + \|\pi_{\beta_2}(y) - \pi_{\beta_2}(x)\|_2 + \ldots + \|\pi_{\beta_k}(y) - \pi_{\beta_k}(x)\|_2
\tag{1}
$$

where each $\|\pi_{\beta_i}(y) - \pi_{\beta_i}(x)\|_2$ is defined as the $\ell^2$ norm for calculating the distance between two anchors regarding the same modality. Note here, that we are only able to compare percepts with the same modalities, such as pairs of visual percepts or pairs of spatial percepts.

### 3.3 Semantic System

We define the semantic system $\Sigma$ as $\langle \Omega, KB \rangle$, which includes the semantic knowledge hierarchy (ontology) $\Omega$, the knowledge base ($KB$), and the corresponding inference and knowledge management mechanisms. The semantic system contains a collection of abstract hierarchically structured knowledge, expressed in some form of formal logic, and includes the set of concepts, instances, relations and rules be-
tween concepts mainly in the domain of commonsense knowledge. Then we can define the semantic description.

**Definition 2 (Semantic Description).** We define the semantic description of an object \( x \in O \) the collection of all grounded instances from concepts in \( \Omega \) as \( \sigma_B(x) = (\omega_{p_1}(x), \omega_{p_2}(x), \ldots, \omega_{p_k}(x)) \).

Then, the semantic distance \( d_\sigma \) between two semantic descriptions regarding two objects \( x \) and \( y \), with \( x \) and \( y \in O \), is defined as:

\[
d_\sigma(y - x) = \| \sigma_B(y) - \sigma_B(x) \|
= \| (\omega_{p_1}(y) - \omega_{p_1}(x)) \| + \| (\omega_{p_2}(y) - \omega_{p_2}(x)) \| + \ldots
\]

\[
\ldots + \| (\omega_{p_k}(y) - \omega_{p_k}(x)) \|
\]

where each \( \| (\omega_{p_k}(y) - \omega_{p_k}(x)) \|_2 \) is defined as the semantic distance metric between the corresponding concepts with respect to the same modality. The semantic distance models how far are two concepts with respect to their semantic content. For example this can be achieved by defining topological similarity distances using the ontology to define the distance between two concepts, or by using statistical means such as a vector space model to correlate concepts in the ontology tree.

### 3.4 Grounding Relations

We define the semantic grounding relations \( G \), as the processes which match the semantic descriptions \( \sigma \) with perceptual signatures \( \pi \). Every percept \( \pi_B(x) \) in the perceptual signature \( \pi_B(x) \) is mapped to a set of concepts or instances \( \omega_B(x) \in \Omega \) which is the corresponding ontological knowledge fragment of the percept.

**Definition 3 (Semantic Grounding Relation).** We define the semantic grounding relation \( G \) which grounds the percept \( \pi \) of a modality \( \beta \) regarding an object \( x \in O \) as \( g_{\beta} \subseteq \Omega \times \beta \times R \in KB \).

The semantic grounding relations are twofold processes. Initially they ground the different percepts to their mapped symbols denoting the corresponding concepts, while then they associate and validate the sets of symbols against the concepts, instances; direct relations and generalizations from the ontology \( \Omega \) of the \( KB \) that represent the modality and the grounded knowledge. The duality of the grounding relations (symbolic and semantic) is necessary in order to ensure that not only we are grounding the percept to the correct symbol, but also that this symbol corresponds to the appropriate concept in the knowledge base (KB). It is important to mention that the semantic grounding relations \( G \) concern object properties, however anchoring concerns objects.
3.5 Anchor

After we have specified the perceptual and semantic systems as well as the semantic grounding relations we can describe the correspondences between percepts and the grounded concepts regarding an object $x$ from $O$ in the environment $E$. These correspondences are reified in an internal data structure which is called anchor and is denoted as $\alpha(x)$. Hence the anchor regarding an object $x$ at time $t$ (i.e. $\alpha(x,t)$) can be defined as any partial function from time to triples in:

$$\alpha(x,t) : \Omega \times B \times F \mapsto \mathbb{R} = \{uid, \pi_B(x,t), \sigma_B(x,t)\}$$

At every moment $t$, an anchor $\alpha(x,t)$ contains three elements: i) a unique symbol $uid$ meant to denote the object $x$ in the anchoring space; ii) the perceptual signature $\pi_B(x,t)$; and iii) the corresponding semantic description $\sigma_B$ of the object (see Fig. 3 for an example). The anchoring process is responsible for creating and maintaining anchors when either new percepts are generated from the perceptual system or when concepts from the semantic system need to be grounded to percepts. Therefore the anchor links the perceptual system $\Pi$ with the semantic system $\Sigma$.

One anchor $\alpha$ is grounded at time $t$, if it contains the percepts perceived at $t$ and the current descriptions. If the object is not observable at $t$ and so the anchor is ungrounded, then the ‘null’ percept $\emptyset$ is stored into the anchor. However $\alpha(x,t)$ still contains the best available estimate since the last observation with respect to the semantic description.

We can then represent the anchoring space of an agent, by $A_{ag} = \{\alpha(x_1), \alpha(x_2), \ldots, \alpha(x_i)\} \forall x_i \in O$. It is essentially a multi-modal space where heterogeneous items of information are mapped, and represents the combined past and current perceptual and conceptual state of an agent, described in terms of perceptual signatures and their corresponding semantic descriptions.
Here we have to mention that grounded anchors represent currently perceived objects, while ungrounded anchors are not deleted from the anchoring space but since they are ungrounded they are considered as past perceptual experience and therefore their content is kept in the anchoring space as well as the KB.

3.5.1 Anchor Indiscernibility Relation

Let $\alpha(x), \alpha(y)$ be two anchors from $A$ regarding two objects $x$ and $y \in O$. The anchor indiscernibility relation $\sim$ for one modality $\beta$ is defined as:

$$\sim_\beta = \{(x, y) \in A \times A : \|\alpha_\beta(y) - \alpha_\beta(x)\|_2 = 0\}$$

(4)

where $\|\alpha_\beta(y) - \alpha_\beta(x)\|_2$ represents the distance between two anchors. Since an anchor is a composite structure with two components which have distance measures separately, from Eq. [1] and [2] we can obtain:

$$\sim_\beta = \{(x, y) \in A \times A : \|d_x^\pi + d_y^\sigma\|_2 = 0\}$$

(5)

$d^*$ denotes that the corresponding distance may or may not be present as for example in the case where we are matching against an anchor that does not have a perceptual signature part and therefore we are not able to compute the distance to the perceptual signature of another anchor. Using the anchor indiscernibility relation, objects with matching signatures and descriptions can be grouped to form classes called elementary sets defined by:

$$C_{\sim_\beta} = \{o \in O | o \sim_\beta c \forall c \in C_{\sim_\beta}\}$$

(6)

3.5.2 Anchor Tolerance

Most of the time we want to facilitate observations of associations in the anchoring space $A$. We can introduce some tolerance value $\varepsilon$ ($\varepsilon \in \mathbb{R}$) so as to obtain the anchor tolerance relation $\cong_{\beta, \varepsilon}$ or simply:

$$\cong_{\beta} = \{(x, y) \in A \times A : \|\alpha(y) - \alpha(x)\|_2 \leq \varepsilon\}$$

(7)

The tolerance relation $\cong_{\beta, \varepsilon}$ is a generalization of the indiscernibility relation, which is easy to see, if in Eq. [5] we set the tolerance value $\varepsilon = 0$.

3.5.3 Anchor Nearness Measure

The anchors $\alpha(x)$ and $\alpha(y)$ within $A$ are considered weakly near each other with the following weak nearness relation.
\[ \alpha(x) \bowtie_{\beta} \alpha(y) \iff \left( \exists x \in \alpha(x), y \in \alpha(y), \exists \beta \in B : (x \bowtie_{\beta} y) \right) \quad (8) \]

While, an anchor \( \alpha(x) \) is considered near to another anchor \( \alpha(y) \) when:

\[ \alpha(x) \bowtie_{\beta} \alpha(y) \iff \left( \exists f_{1}, f_{2} \subseteq F, \exists f \in F, \exists A \in O_{/\sim_{f_{1}}}, \exists B \in O_{/\sim_{f_{2}}}, \exists C \in O_{/\sim_{f}} : (A \subseteq \alpha(x)) \land (B \subseteq \alpha(y)) \land (A, B \subseteq C) \right) \quad (9) \]

With the Nearness Measure \( NM \) we want to measure the similarity of the “things” two weakly near anchors \( \alpha(x), \alpha(y) \) have in common. Let \( \alpha(z) = \alpha(x) \cup \alpha(y) \). By \( [\alpha(z) /_{\beta} \alpha(x)]_{\alpha(x)} = \{ \alpha(z) \in \alpha(z) /_{\beta} | \alpha(z) \in \alpha(x) \} \), we denote the portion of the tolerance class \( \alpha(z) /_{\beta} \) that belongs to \( \alpha(x) \), and similarly \( [\alpha(z) /_{\beta} \alpha(y)]_{\alpha(y)} = \{ \alpha(z) \in \alpha(z) /_{\beta} | \alpha(z) \in \alpha(y) \} \) to denote the portion of \( \alpha(y) \). Then using Eq. 8 let \( \alpha(x) \) and \( \alpha(y) \) be weakly near and \( \alpha(z) /_{\beta} \) a covering of \( \alpha(z) \) defined by \( \bowtie_{\beta} \). Then a \( NM \) is formulated in (Eq. 10).

\[
NM_{\bowtie_{\beta}}(\alpha(x), \alpha(y)) = \left( \sum_{\alpha(z) /_{\beta} \in \alpha(z) /_{\beta} \bowtie_{\beta}} |\alpha(z) /_{\beta}| \right)^{-1} \]

\[
\cdots \sum_{\alpha(z) /_{\beta} \in \alpha(z) /_{\beta} \bowtie_{\beta}} \min(\frac{|[\alpha(z) /_{\beta}]_{\alpha(x)}|}{\max(\frac{|[\alpha(z) /_{\beta}]_{\alpha(x)}|}{|[\alpha(z) /_{\beta}]_{\alpha(y)}|})})
\]

The idea behind the \( NM \) is that sets that are similar should have a similar number of objects in each tolerance class. Thus, for each tolerance class obtained from the covering of \( \alpha(z) \), the \( NM \) counts the number of objects that belong to \( \alpha(x) \) and \( \alpha(y) \) and takes the ratio (as a proper fraction) of their cardinalities. Furthermore, each ratio is weighted by the total size of the tolerance class (thus giving importance to the larger classes) and the final result is normalized by dividing by the sum of all the cardinalities. The range of the \( NM \) is in the interval \([0, 1]\), where a value of 1 is obtained if the sets are equivalent and a value of 0 is obtained if they have no elements in common.

### 3.6 Anchor Management

To support the proposed framework we further describe the different functionalities across all the anchoring framework which are enabling the support of both top-down and bottom-up approaches [37]. Essentially the terms “top-down” and “bottom-up” represent the two ways of traversing within the anchoring space, semantically and perceptually respectively. “Bottom-up” information acquisition is driven by an event
originating from a sensing resource (e.g. the recognition of an object in an image) as soon as there are new percepts generated in the perceptual system. This leads to the creation of an anchor and the instantiation of the corresponding concepts and instances and their relations in the semantic system.

In “top-down” acquisition, a concept or logical formula from the semantic system is anchored to percepts on request, for instance originating from a planner; a reasoner, or a natural language interface. This is done by traversing downward into the anchoring space and where possible, converging into an anchor that contains the corresponding percepts and concepts which are compatible with the original formula. Therefore the different functionalities in order to utilise the top-down and bottom-up information management are namely, (Re)Acquire and Augment functionalities for bottom-up acquisition and for top-down retrieval we have the Query and Find functionalities. All these functions end up relying on the Track and Match functions which internally use the Nearness Measure described in the previous section.

3.6.1 Bottom-up Information Management

(Re)Acquire initiates a new anchor whenever a percept is received, which currently does not match any existing anchor. It takes a percept $\pi(x)$, and returns an anchor $\alpha(x,t)$ defined at $t$ in the anchoring space $A$. In bottom-up acquisition, a randomly generated symbol (i.e. the Unique Identifier $uid$) is attributed to the anchor. Further-
more, information about the object, its properties and relations are included into the anchor via the semantic grounding relations.

Augment, at each perceptual cycle, loops through the semantic descriptions of the anchoring space. It tracks changes in the anchors and appends the logical sentences about the agent’s perceptions into the semantic system. It can be thought as a “semantic track” which essentially keeps the perceptual knowledge in the knowledge representation system coherent with the perceptual signatures of the observed objects. While at first this function appears trivial, synchronizing the concepts, instances and relations of the objects in the anchoring space with the ones in the knowledge base, requires further verification so as to avoid conflicts or unrelated assertions.

3.6.2 Top-down Information Management

Find Takes a semantic description $\sigma$ and returns an anchor $\alpha(x,t)$. It evaluates if existing anchors that have already been created by the Acquire satisfy the semantic description or corresponding perceptual signature, and in that case, it returns the matches. Otherwise, it performs a similar check with existing percepts (in case, the description does not satisfy the constraint of percepts considered by the Acquire. If a matching percept is found an anchor is retrieved.

Query Accepts a logical formula originating from an external component such as a planner or a natural language interface, that needs to be inferred from perceptual observations and utilizes the semantic system to produce the satisfiability of the formula according to the perceptual observations. Query can be understood better if we describe it as the semantic “Find”. While Find accepts a symbolic description and converges into anchors that contained the compatible percepts, Query accepts a logical formula, and converges into grounded concepts from the anchors in the anchoring space. For instance, this logical formula could be a precondition axiom of a planner that needs to be satisfied in order for a plan to be executed.

3.6.3 Tracking and Matching

Track The track functionality takes an anchor $\alpha(x,t)$ defined for $t - k$ and extends its definition to $t$. The track assures that the percept pointed to by the anchor is the most recent and adequate perceptual representation of the object, while ensuring that the corresponding grounded semantic description reflects the current state of the perceptual signature. We consider that the signatures and descriptions can be updated as well as replaced but by preserving the anchor structure over time we affirm the persistence of the object so that it can be used even when the object is not currently perceived (caused by the object being out of view and/or by the errors generated during the measurement of perceptual data). This facilitates the maintenance of information while the agent (i.e. robot) is moving. It also supports
stable representations of the world at a symbolic level, in the long term and without carrying perceptual glitches.

**Match** The match function is the core, multi-modal function, which may accept a perceptual signature or a semantic description or both, and it returns the anchors which match the required specification(s), by matching all the knowledge fragments (if any) against the semantic descriptions of the anchoring space, and by matching all the percepts (if any) against the perceptual signatures. The matching relies on the *NM* between two anchors.

### 4 Studies using the Anchoring Framework

In the following sections we present several case studies and investigations in which we have applied the augmented anchoring framework described the previous section. Specifically we are interested in the challenging problem of deriving high-level descriptions of visual information from multiple sensing modalities. We also consider other important aspects, such as the applicability of the model in realistic settings. Anchoring is used for grounding the interpretations of the semantic representations in perceptual information originating from the environment as well as other information sources. Further, we elaborate on the integration of knowledge representation and perceptual techniques. We explore this integration for the purpose of fusing commonsense knowledge in the anchoring process. Then we show how anchoring is done, when using spatial relations and also between heterogeneous perceptual agents. Finally we discuss some preliminary results on anchoring conceptual knowledge when using information from non physical sources, such as the web.

### 5 Anchoring with spatial relations

In \[46\] we used symbolic knowledge representation and reasoning capabilities to enrich perceptual anchoring. The use of the *KR&R* is advocated to allow the human to assist the robot in simple anchoring tasks, such as the disambiguation of objects, thereby exploring a deeper form of mutual human robot interaction (*HRI*). The knowledge base consists of two parts: *i* the terminological component (*T*Box), that contains the description of the relevant concepts and their relations; and *ii* an assertional component, the *A*Box for storing concept instances and assertions on those instances. For the anchoring domain we require an ontology that covers all the physical entities and the corresponding perceptual properties, which are recognised by the perceptual system and eventually occur during an anchoring scenario.

In addition we used knowledge that was inferred from basic knowledge contained in the anchors and also collected from external sources using other cognitive capabilities, such as other anchoring processes or linguistic interaction. Modelling an
ontology is in general a difficult task, therefore in this particular instance we adopt
an ontology based on a subset of the ontological framework DOLCE (A Descriptive
Ontology for Linguistic and Cognitive Engineering) [39], an upper-level ontology
developed for the Semantic Web. The knowledge base is implemented using the
LOOM knowledge representation system [5].

5.1 Semantic modelling of spatial relations

Spatial relations were used in the symbolic description of objects and allowed to dis-
tinguish objects by their location with respect to other objects. They also played an
important role when it came to HRI. Two classes of binary spatial relations between
a reference object and a located object were considered: a) the topological (distance)
relations “at” and “near”; and b) the projective (directional) relations “in-front of”,
“behind”, “right”, and “left”. The interpretation of a projective relation depends on
a frame of reference; for simplicity we assume a deictic frame of reference with an
egocentric origin coinciding with the robot platform. We finally modelled the spatial
relations as concepts in the ontology, where we considered each spatial relation of
the sub-concept of “Abstract Relation”, to have as properties a reference object, a
located object, and a spatial region. Using the LOOM KR&R system we were able
to perform some basic inferences with the information provided by the anchoring
component. Thus, we could focus on the high-level aspects behind the logical rep-
resentation and the dynamic properties of the objects, instead of invoking directly
sensory information. The user interface was based on a plain text-based application,
where the user can type in sentences in simple English. The sentence is analysed by
a recursive descent parser and translated into a symbolic description.

The grammar allows commands of the form FIND . . . followed by
the description of the object. The description consists of a main part
that can be followed by sub-clauses
describing objects that are spatially
related to that object.

The main part and each of the
sub-clauses can be either a defi-
nite or indefinite description, indi-
cated by the article “a” or “the”. It
also includes the object’s class, for
eexample “cup”, and optionally its
colour and smell. The smell of an
object is inferred from the clause
“with . . .” following the object’s
class, indicating that it contains a
liquid; for example, “the cup with

Fig. 5 Subset of the LOOM ontology used in the de-
velopment of the knowledge representation layer of
the implemented system. From [36].
“coffee” is assumed to be a cup containing coffee, and as such, smelling of coffee. The derived symbolic description is used to construct a query for the KB. The main functionality is realised by the FIND routine, which collects candidates from the KB that match the description. If there is more than one candidate, the anchoring module evaluates further the properties in the given description, apart from shape and colour, and selects those additional properties. If an ambiguity still persists the system asks the user about which of the candidate objects to select.

### 5.2 System validation

The validation of the system was done in the context of an intelligent home environment, which was used for ambient assisted living for the elderly or disabled. In this smart home, an existing middle-ware framework called the PEIS-Ecology [45] is used in order to coordinate the exchange of information between the robot, other pervasive technologies in the environment and the human user. To illustrate the utility of KR&R in anchoring three case studies were presented.

The first focussed on the inclusion of spatial relations in the anchoring framework. The set of binary spatial relations mentioned in the previous section were used in the validation. As spatial prepositions are inherently rather vague, we are using fuzzy sets to define granular spatial relations. The proposed method computes a spatial relations-network for anchored symbols and stores that in the KB. The robot surveys a static scene with three objects (two green garbage cans and a red ball) and the anchoring module creates anchors for these objects in a bottom-up way (as soon as new percepts are identified by the vision system).

We examined the possibility to use spatial relations to query for an object, for example: “find the green garbage to the left of ball”. Similarly, a human user can be asked to resolve an ambiguity in a find request. For instance the query “Find the green garbage” returns more than one anchor. Therefore the anchoring module determines an anchored object that is spatially related to these anchors as reference object and presents the user with a choice, enumerating the returned anchors and their spatial relation(s) to the reference object. Then the query is reformulated using also the selected relation(s).

A second case used multi-modal information about objects, including both spatial information and in this case, olfactory information given by an electronic nose (See Fig 6). The KR&R system is then used to assist the robot in determining which perceptual actions can be taken to collect further information about the properties of the object. As in the previous case, where an ambiguity is introduced, the robotic system uses its olfactory module for resolving an ambiguous case among different cups. The robot is given the command “Find a green cup with coffee”. Four candidates were found in the KB, which were asserted to be cups, i.e, vessels containing liquid with an associated smell. This triggers the task planner to generate a plan to visit each candidate and collect an odour property. The ambiguity is then resolved once a first match is found.
The third case investigated the possibility to perform reasoning about object properties in order to determine an optimal candidate. We considered the case where there were some scattered fruits on the floor. The robot navigated around the floor and correctly perceived; classified and asserted the instances into the KB. We then asked the robot to pick the orange that was to the left of the apple (spatial misinformation). Since this was not the case, it responded with a valid proposition, that there is no orange to the left of the specific apple and instead there is one on the right of the banana. The user confirms that the alternative candidate is indeed the requested object. As a next step we asked the robot to pick a banana to the right of the apple. Here there are 2 candidates, both being to the right of the apple. However one of them being classified as rotten and therefore not edible. The system informs the user about the situation and after removing this option, it prompts that finally there was one banana on the right of the apple (suitable for consumption) and returns.

6 Commonsense Knowledge and Anchoring

In [18 21] we have investigated the integration of a large KR&R with a perceptual system which consisted of networked sensors. The role of the KR&R system is to maintain the world model which consists of the collection of the semantic information perceived by the robotic assistant and a smart home. The commonsense KR&R system used in this work is Cyc [34]. Operations, such as assertions, retractions, modifications and queries are stated using Cyc’s formal language, CYCL. Such a system, that rests on a large scale general purpose knowledge-base, can potentially manage tasks that require world knowledge (or common-sense).
6.1 Knowledge Integration and Synchronisation

To enable synchronisation between the KR&R system and the anchoring system three aspects were considered: \(a\) defining anchoring in the KB; \(b\) handling of assertions; and \(c\) handling of ambiguities. CYC is not consistent globally but rather attempts 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 commonsense reasoning about this knowledge. For instance, if the location of the anchor of the object “cup” is kitchen, then the object and the kitchen are instantiated into CYC, inheriting all the properties related to the generalised concepts Kitchen and Cup, such as ManMadeThing.

To synchronise perceptual data with knowledge in CYC, anchored objects from both working and archive memories were instantiated. Then, the knowledge manager keeps 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 represent the agent’s perception of the object (See Fig. 7). When a particular predicate of the symbolic description changes (e.g. the location of an object), an update of the corresponding predicate asserts the new information into the KB.
6.2 Knowledge Management

The main challenge of integrating a large KR&R like CYC is to be able to synchronise the information with the perceptual data coming from multiple sensors, which is inherently subject to incompleteness, glitches, and errors. The anchoring system provides a stable symbolic information despite fluctuating quantitative data. 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.

Ambiguities eventually arise during the synchronisation process and they 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 grounded concept using candidates from the KB which denote the specific eventual concept.

Another aspect we consider in the synchronisation process, is the information about the visibility of the object. This is important as the agent still needs to maintain the relevant knowledge about a recently perceived object, despite the fact that the corresponding perceptual information may not currently be available to the agent. For example the robot still needs to maintain the knowledge that “a cup is located in the kitchen”, even though the “cup” might be out of its current field of view (epistemic knowledge). This was achieved by 2nd order rule assertions.

6.3 Evaluation

As part of the test-bed we used a physical facility, which looks like a typical bachelor apartment of about 25m². It consists of a living room, a bedroom and a small kitchen. We initially presented the robot with 15 objects, while capturing 2 to 5 training images per object from different viewpoints. It is important to mention that the robot recognised instances of objects and not object categories. We placed those objects around the smart 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 recognise and anchor those objects. Finally a human communicated with the system in natural language, by typing into the console based graphical user interface, questions (related to the perception of the robot as well as commonsense), or communication commands.

Excerpts of dialogues that reflect perceptual information from the mobile robot and the environment are presented in Table 1. We were also able to use the visual sensors in the ceiling to observe things that either the robot could not perceive or were occluded. Another instance of the object recognition component was trained on large objects from the smart home (e.g. television set, couch, table). Through the knowledge base assertions, which contained a larger collection of objects from the environment, the robot was able to compensate for its limited perception. This aspect
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 1 Excerpt from the dialogue between the human and the Anchoring Robot. From [21].

spawned an interesting problem of how to coordinate this cooperative perception with respect to the semantic knowledge base, in the context of anchoring. In sum, clearly the advantage of using a large scale deterministic 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 also be useful when queries about functions and properties of objects are made. Due to the expressiveness of CYC we could differentiate perceptual and epistemic knowledge, at the same time keeping both consistent. Using linguistic interaction we could study how anchors were formed, while probing the symbolic content.

7 Cooperative Anchoring using Perceptual & Semantic Knowledge

In a third instance [19] we introduce a model for semantic cooperative anchoring, with which we intend to capture the underlying processes for systematically transforming collectively acquired perceptual information into semantically expressed commonsense knowledge, over a number of different agents. We enable the different robotic systems to work in a cooperative manner, by acquiring perceptual information and contributing to the collective semantic KB. We adopt the notion of the local and global anchoring spaces from [32]. The different perceptual agents maintain their local anchoring spaces, while the global anchoring space is coordinated by a centralised component which hosts also the commonsense KB. The reasons behind the centralised module concern the computational complexity of the commonsense KB (CYC).
7.1 Semantic integration

This integration where we use commonsense information, is capable of supporting better (closer to linguistic) expressivity in human robot communication and generally improves the knowledge and reasoning abilities of the agent. It provides an abstract semantic layer that forms the base for establishing a common language and semantics (i.e. ontology) to be used among robotic agents and humans. **Semantic integration** takes place at two levels, in accordance with the design of the anchoring framework.

We have developed a local $KB$ which is part of the anchoring process of every agent. Our solution involves a local knowledge representation layer which aims at a higher degree of autonomy and portability. More specifically, it: $a$) is a lightweight component; $b$) adheres to the formally defined semantics of the global $KB$ (CYC); $c$) provides simple and fast inference; $d$) allows communication between other components (e.g. the global anchoring agent; CYC or the Semantic Web); and $e$) is based on the same principles that our core knowledge representation system (CYC) is based.

Each local $KB$ is a subset of a global $KB$ which is managed by the global anchoring agent that coordinates the different local anchoring processes. The global knowledge representation is used as the central conceptual database and $a$) defines the perceptual context of agents; $b$) defines the formal semantics; $c$) enables more complex reasoning; and $d$) contains the full spectrum of the commonsense knowledge in the $KB$. The duality of knowledge representation systems is due to the fact that we want to enable semantic knowledge and reasoning capabilities on each perceptual agent, however we cannot afford (yet) to embed a complete commonsense knowledge base instance on every agent.
7.2 Implementation

We have implemented our approach using the PEIS distributed framework [45] with the application domain being a smart home for the elderly. The implementation involved the mobile robot which is capable of perceiving the world through standard heterogeneous sensors and is acting as the mobile anchoring agent. An ambient anchoring agent pertained sensors and actuators embedded in the environment, such as observation cameras, localization systems. Finally, a high-end desktop computer undertakes all the computationally demanding tasks related to the commonsense knowledge base.

The evaluation of the system includes how well the system performs, when recognising, anchoring and maintaining in time the anchors referring to the objects in the environment. We have tested several scenarios, where in particular we considered the case in which the knowledge from the perceptual anchoring agents populating the CYC knowledge-base was used to enable inference upon the instantiated perceptual concepts and sentences. It was possible to use the object’s category, colour, spatial or topological relations, along with other grounded perceptual information, in order to a) refer to objects; b) query about objects; and c) infer new knowledge about objects. In addition, we have considered the use of commonsense knowledge in different contexts, through the multi-context inference capabilities of the CYC system.

In order to monitor and allow interaction with the anchoring framework, we have developed an interface which has three ways for providing access to information. It
includes a primitive natural language interface to interact with the different agents. It also displays the state of the system and provides runtime information about the performance of anchoring. The main component and observation feature of the tool is the visualisation of the 3D augmented virtual environment. It is simply, a scene of the environment rendered in a 3D engine which encapsulates every non-perceptual object from the environment, such as the wall surroundings. Then information acquired from the global anchoring space was used to virtually reconstruct the perception of each agent in the rendered scene-graph.

Finally we have shown how the different perceptual agents can communicate using logical sentences, in order to exchange information between each other, in a cooperative manner. We focused on the locally maintained knowledge of each perceptual agent in the context of querying other available agents to complement for the limited perceptual knowledge so as to finally infer the answers to the queries.

8 Towards Concept Anchoring using information from the World Wide Web

Finally we consider that autonomous robots which interact with humans, require open ended systems where domain knowledge can be dynamically acquired and is not necessarily defined a-priori. The World Wide Web (WWW) is a viable source of knowledge that would satisfy the requirement of being available on demand while it
evolves continuously, with contributions from human users. Hence, we explored the possibility of integrating the knowledge coming from a non physical environment such as the WWW, in the anchoring process. This was achieved using an augmentation of the anchoring framework, called Conceptual Anchoring [20].

8.1 Representation

Acquired data, from a non-physical information source, specifically the web were used so as to form representations (conceptual anchors), about concepts that might not necessarily exist in the agent’s real environment. Via the conceptual anchoring space, the agent was able to: a) express concepts in terms of their associated (eventual) percepts and symbolic knowledge; and b) use the information from the web, in order to guide the learning of novel perceptual instances (physical objects). In the context of conceptual anchoring, the percepts produced from on-line resources were used for the construction of the perceptual signatures. The latter were grounded to their corresponding semantic descriptions and tangible semantic knowledge from the commonsense KB, in a similar way as in perceptual anchoring. The multi-modal data structure (conceptual anchor) holds the perceptual data, semantic descriptions and unique identifier about one conceptual object.
Since the conceptual percepts do not correspond necessarily to physical entities in the agent’s environment, implicitly the conceptual anchor does not correspond to a specific physical object (perceptual instance), but rather to the generalised conceptual appearance and implicit knowledge about the general concept. For instance the conceptual anchor of the concept “Cup” links the multiple visual percepts of cup images and features of them, with the corresponding grounded visual concepts such as the colours or shapes of cups and with the tangible semantic knowledge about cups like: Cup is a specialization of DrinkingVessel, inheriting all the implicit relations that it is also an Artifact, a SpatiallyDisjointObject, a Container, … etc. Since both conceptual and perceptual representations are expressed as anchors, the anchoring functionalities are able to handle both representations in the same way, especially regarding matching.

8.2 Preliminary results

Conceptual Anchoring is a system built on top of the knowledge based perceptual anchoring architecture and mainly a) provides an interface to acquire perceptual and semantic knowledge from the web, while b) maintains the conceptual anchors in the anchoring space which are used to store the conceptual models. By acquiring conceptual anchors, which can be thought as “meta-anchors”, we were able to use the multi-modal conceptual representations, as perceptual templates which guided the detection and learning of existing physical objects (perceptual instances) in the environment, without having previously trained the robot to recognise each instance individually (a tedious and highly technical process).

So far the notion of the conceptual anchors has been tested in a proof-of-concept scenario and focus of the evaluation has been on the ability to retrieve the conceptual anchors from the web using keywords. While conceptual anchoring is the recent development in the anchoring framework, still needs to be further enhanced as it is believed that relying more on dynamic information from the WWW will be the dominant trend in the field.

9 Discussion

In order to realise an anchoring framework with such a dynamic nature that it can fully integrate perception and knowledge on demand, it is necessary from a theoretical point of view to relax an important assumption of current AI systems, namely the closed-world assumption. Specifically, current anchoring frameworks operate in dynamic environments where objects are moved and new instances of objects are added or removed, but it is still assumed that the different classes of objects that are of interest are known and trained a-priori.
This instance of the closed-world assumption, presupposes that all aspects of the world (be they the categories of interesting objects in anchoring, or a predicate modelling of a fact in planning) that are not known to be true, are considered as false. In addition, the decision to create a new instance of an object in the knowledge base or to assign the current perceptual information to a known object is currently based on information about appearance and location. Abandoning the closed-world assumption would allow to handle objects whose category is not defined a-priori, as well as to handle new knowledge that can be discovered dynamically.

We require access to both general conceptual knowledge about objects and the specific knowledge about the environment in which the system operates. The general knowledge is necessary to establish and give content to new categories while the specific knowledge helps in uniquely identifying objects in the environments to prevent a proliferation of possible instances of objects and facilitate the reasoning about objects and their use.

Eventual solutions can be built upon existing components, such as activity recognition, planning, advanced perceptual routines and multiple sensors, all of them interacting with an anchoring module. The anchoring process is able to collect and fuse information from heterogeneous sensors while grounding percepts to the corresponding concepts, before making appropriate assertions of the instantiated knowledge in a knowledge base.

This chapter has outlined the main contributions regarding perceptual anchoring and its development towards including more complex and dynamic semantic information. In particular, we have considered how anchoring can interact with a large scale deterministic commonsense knowledge-base using perceptual information regarding physical objects. Using the basic building blocks of the anchoring framework as described, we have shown how anchors can be used in the management of perceptual and semantic information. The challenge for anchoring in the future will be to determine how other types of information, such as activities or object use, can also be included into the framework.

References


