A Hybrid Cognitive System for Spatial Reasoning in a Robotics Framework

Alessandro Oltramari AOLTRAMA@ANDREW.CMU.EDU
Jerry Vinokurov JERRYV@CMU.EDU
Christian Lebiere CL@CMU.EDU
Functional Modeling Lab, Psychology Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Anthony Stentz TONY@ANDREW.CMU.EDU
Jean Oh JEANOH@REC.RI.CMU.EDU
Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Abstract

In this paper we present a hybrid cognitive system to support robot navigation. The system leverages the ACT-R cognitive architecture by integrating an external knowledge component for high-level reasoning. Methodological and implementation aspects are provided, together with an overview of the system's functionalities and the results of a preliminary simulation experiment.

1. Introduction

Humans constantly negotiate space with a variety of entities in the world, moving towards targets, around obstacles, away from threats, sorting perceptual cues according to context, selecting and operating objects that functionally suit intended goals. This complexity of human behavior in the environment gradually emerges as the mind learns to couple visual perceptions to background knowledge and to perform patter recognition and context-driven reasoning. Though no ultimate explanation has been given of how humans can generalize over perceptual contents and create mental representations at the basis of spatial reasoning (Casati & Varzi, 1996), some principles have been distilled. For instance, according to (Tversky, 1977) & (Biederman, 1987), *i)* similar objects have a high degree of overlapping components (scissors and knives, chairs and tables, etc.); *ii)* spatial proximity depends on the context (nearby a hospital, there are probably a parking lot and a heliport); *iii)* objects that serve related ends typically appear together (pens and notebooks are usually found within a short range, since they are both related to the activity of "writing").

These generic principles embed just a small fraction of the large amount of common-sense spatial knowledge humans exploit in accomplishing everyday tasks. But, this knowledge is not provided *a priori*: it is learned by accumulation of experiences. Envisaging the future on the basis of past experiences is one of the primary activities of the human mind: this predictive competence, far too complex to be innate, is *de facto* realized as a mechanism of continuous 'pattern recognition'

(Kurzweil, 2012), where perceptual information and world knowledge come together. In this context, investigating the relationships between sensorimotor stimuli, visual perceptions, spatial knowledge structures and context-driven reasoning, is a key element to understand the mechanisms of high-level cognition that determine human intelligent behavior in space, and that, ultimately can be replicated by artificial intelligent systems. In this regard, the main challenge it's to *situate* artificial systems in a real environment: several attempts have been made to reproduce cognitive aspects in robotic systems (Cassimatis, Trafton, Bugajska, & Schultz, 2004; Kelly & Avery, 2010; Trafton, Hiatt, Harrison, & Khemlani, 2012; Hawes & Klenk, 2012; etc.): "Cognitive Robotics" thus emerged as an interdisciplinary field where AI and Cognitive Science coalesce to approach general problem-solving from a "bounded rationality" perspective (Simon, 1991), rather than relying on the search for optimal algorithmic solutions (Kurup & Lebiere, 2012). If the problem of implementing intelligent systems for navigation and recognition in space goes back to the dawn of robotics (Moravec, 1988), the novelty of the cognitive approach is to take inspiration from human-like solutions.

In this paper we follow this comprehensive approach: in particular, we illustrate an ontology-driven cognitive system for predicting spatial configurations of buildings on the basis of visual cues about their constituent walls and surroundings. In the paradigmatic scenario, a robot navigates through an outdoor environment carrying out tactical behaviors typically issued by commanding officers (e.g. "screen the backdoor of the hotel"): the cognitive system, accordingly, aims at supporting navigation and planning routines with high-level pattern recognition, mimicking the human capabilities of discriminating spatial entities and their features and assembling a mental picture of the surroundings.

2. A Hybrid System for Knowledge-based Cognitive Processing

In this section we briefly introduce the two components of the above-mentioned cognitive system, namely the ACT-R architecture (Anderson, 2007) and the SCONE Knowledge Base Systems (Fahlman, 2006).

2.1 Replicating Cognitive Mechanisms with ACT-R

Cognitive architectures attempt to capture at the computational level the invariant mechanisms of human cognition, including those underlying the functions of control, learning, memory, adaptivity, perception, decision-making, and action. ACT-R (Anderson & Lebiere, 1998) is a modular architecture including perceptual, motor and declarative memory components, synchronized by a procedural module through limited capacity buffers (Figure 1). Declarative memory module (DM) plays an important role in the ACT-R system. At the symbolic level, ACT-R agents perform two major operations on DM: 1) accumulating knowledge chunks learned from internal operations or from interacting with objects and other agents populating the environment and 2) retrieving chunks that provide needed information. ACT-R distinguishes declarative knowledge from procedural knowledge, the latter being conceived as a set of procedures (production rules or "productions") which coordinate information processing between its various modules (Anderson & Lebiere, 1998): accordingly, agents accomplish their goals on the basis of declarative representations elaborated through procedural steps (in the form of *if-then* clauses). This dissociation between declarative and procedural knowledge is grounded in experimental cognitive psychology; major studies in cognitive neuroscience also indicate a specific role of the hippocampus in forming permanent declarative memories and of the basal ganglia in production processes (see Anderson, 2007, pp. 96-99, for a general mapping of ACT-R modules and buffers to brain areas and Stocco, Lebiere, & Anderson, 2010 for a detailed neural model of the basal ganglia's role in controlling information flow between cortical regions). ACT-R performs cognitive tasks by combining rules and knowledge: in these regards two core mechanisms are important in the context of this paper: *i) partial matching*, the probability of association between two distinct declarative knowledge chunks, computed on the basis of adequate similarity measures; *ii) spreading of activation*, the phenomenon by which a chunk distributionally activates different declarative patterns. These two basic mechanisms belong to the general sub-symbolic computation underlying chunk activation, which in ACT-R controls the retrieval of declarative knowledge elements by procedural rules.

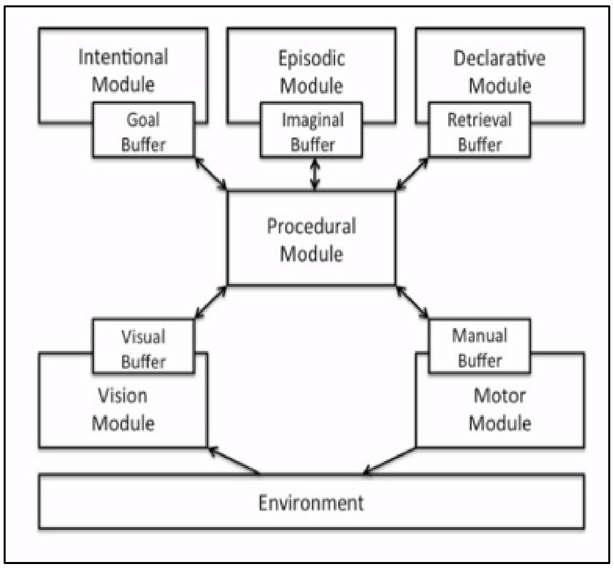

Figure 1 ACT-R Modular Framework

2.2 Common Sense Reasoning through SCONE

Inasmuch as humans make sense of the environment by means of coupling perception with knowledge, the ACT-R cognitive architecture should be enabled to generalize over perceptual inputs by applying fine-grained models of the world to concrete scenarios. In order to fulfill this goal however, ACT-R needs to properly encapsulate those models or *ontologies* (Guarino, 1998) and exploit them for high-level pattern recognition and reasoning. Ontologies are semantic specifications of a given domain or application and are generally used in combination with inference engines for deductive reasoning. Since ACT-R declarative module supports a relatively coarse-grained semantics based on slot-value pairs, and the procedural system is not optimal to effectively manage complex logical constructs, a specific extension is needed to make ACT-R suitable to fulfill knowledge-intensive tasks. Accordingly, we engineered an extra module as a bridging component between the cognitive architecture and an external knowledge-base system, SCONE (Fahlman, 2006). SCONE is an open-source knowledge-base system (KBS) intended for use as a component in many different software applications: it provides a framework to represent and reason over symbolic common-sense knowledge. Unlike most diffuse KBS, SCONE is not based on Description Logics (Staab & Studer, 2004): its inference engine adopts marker-passing algorithms (Fahlman, 2006), originally designed for massive parallel computing, to perform fast queries at the price of losing logical completeness and decidability. In particular, SCONE represents knowledge as a *semantic network* whose nodes are locally weighted (*marked*) and associated to arcs (*wires*[1]) in order to optimize basic reasoning tasks (e.g. class membership, transitivity and inheritance of properties, etc.). The philosophy that inspired SCONE is straightforward: from vision to speech, humans exploit the brain's massive parallelism to fulfill all recognition tasks; if we want to deal with the large amount of knowledge required in common-sense reasoning, we need to rely on a mechanism that is fast and effective enough to simulate parallel search. Shortcomings are not an issue since humans are not perfect inference engines either. Accordingly, SCONE implementation of marker-passing algorithms aims at simulating a pseudo-parallel search by assigning specific marker bits to each knowledge unit. For example, if we want to query an ontology of automotive body design to get all the parts of a car body, SCONE would assign a marker M1 to the ‘A-node’ *Car* and search for all the statements in the knowledge base where M1 is the A-wire (domain) of the relation *part of*, returning all the classes in the range of the relation (also called `B-nodes'). SCONE would finally assign the marker bit M2 to all B-nodes, also retrieving all the inherited subclasses (e.g. *window section*, *rear panel*, *hood vent*, etc.). The implementation of ontologies with SCONE allows for effective formal representation and automatic inferencing of knowledge structures.

2.3 Integrating ACT-R with SCONE

The proposed integration between SCONE and ACT-R abides by the general cognitive constraints of the architecture: in particular, we extended ACT-R with a new module to bridge the architecture and the KBS, keeping their distinctive mechanisms properly separated. The “SCONE module”, as we simply call it, operates in a standard ACT-R fashion, where buffers are used to evaluate chunks when suitable requests are dispatched. SCONE module requests are designed to match legal queries in the KBS, which check the validity of a statement by means of the inference engine[2]: in our approach, this mostly happens when generalization over input signals can help detecting high-level patterns of information[3]. In principle, this is also possible when only partial information is available: partial matching and spreading activation mechanisms can fill the gaps and trigger the retrieval of the best-matching knowledge structures from a given ontology stored in SCONE. In the next section we present more in detail how these functionalities of the cognitive system have been instantiated in a model that predicts the topological and morphological features of buildings on the basis of constituent walls and surrounding features (grass, sidewalks, gravel, road signs, trees, garages, etc.). If it is true that ACT-R can *per se* deal

[1] In general, a *wire* can be conceived as a binary relation whose domain and range are referred to, respectively, as ‘A-node’ and ‘B-node’.
[2] The SCONE marker-passing algorithms are similar to ACT-R spreading activation, leaving open the possibility of a deeper integration of the two frameworks in future work.
[3] Though this is currently the result of a modeler’s design choice rather an independent computational mechanism encapsulated in the architecture.

with simple logical reasoning on the basis its production mechanisms, when knowledge-intensive tasks come into play an external KBS like SCONE becomes a critical plug-in for augmenting ACT-R scalability, computational efficiency, and semantic adequacy.

3. A Cognitive Model for Building Prediction using Background Knowledge

3.1 General Aspects

We present here a hybrid model for predicting the geometry of buildings on the basis of their constituent walls and spatial and contextual knowledge (section 3.3). The model has been built in the framework of the hybrid cognitive system constituted by ACT-R cognitive architecture and SCONE KBS. Section 3.2 outlines the suite of spatial ontologies encoded in SCONE and used by the model to perform high-level reasoning. The robotic architecture also includes a component called the "World Model," which effectively serves as an object store for elements identified by the robot's perceptual systems. In the application scenario, as long as the robot *perceives* building components such as walls, doors, windows etc. and the World Model is populated with the resulting information (which plays the role of central information repository for all the other components), the ACT-R model is used to process the incremental data, and update the shape of a given building according to progressive predictions. In this way, the robot has access to incremental projections of the geometry of a building, which can then be synchronized with the navigation planning process. At this stage, the integration between the model and the RCTA architecture is currently under development. For this reason, we ran a synthetic simulation within the ACT-R environment to evaluate the model's predictions (section 4).

3.2 Modular Ontologies for the Representation of Space

Despite the extensive literature on using spatial reasoning in information systems (Bateman & Farrar, 2005), to our knowledge there has been just one attempt to apply spatial ontologies to tactical behaviors executed by unmanned ground vehicles (BouSaba, Esterline, Homaifar, & Fatehi, 2008). In this respect, our approach aims at making a step further with respect to state-of-the-art, by actually *engineering* and *testing* spatial ontologies into a full-fledged cognitively-driven robotic system. For this purpose we have developed HORUS (Hybrid Ontology for the Representation of Unified Space), a comprehensive ontology of buildings created by mapping primitive concepts and relations extracted from a library of spatial ontologies to the World Model's semantic data types. HORUS is essentially grounded in two open-source ontologies: DOLCE-Lite, which includes basic spatial concepts and related properties (part, region, extension etc., see Borgo & Masolo, 2009) and RCC-Ontology, including primitive topological relations like "connection", "overlap", etc. (Kutz, Lücke, & Mossakowski, 2008). HORUS has the twofold purpose of 1) providing machine-readable semantic specifications of the World Model datatypes and 2) structuring and populating the declarative knowledge of an ACT-R model for building prediction. Regarding point 1), we have mapped the spatial representation of buildings and walls of a test site (Fort Indiantown Gap – FTIG – see Figure 2 for an aerial view of the facility) with geometrical and morphological properties adopted by the navigation planner of the RCTA architecture. Integration was engineered by using the suite of different spatial ontologies mentioned before. In this sense, HORUS exemplifies a modular approach, where conceptual qualitative and quantitative spatial properties are considered as separate information layers from the representational and inferential viewpoint, though being incorporated into a common ontological infrastructure (a quantitative ontological layer for the representation of discrete measures is also available, though in this paper we are mainly focusing on qualitative spatial reasoning).

Figure 3 illustrates the main distinction between architectural and functional features of walls and buildings, where the former are internally constrained by metric and construction characteristics (quantitative layer), while the latter refer to the spatial regions defining the area of operations of the mobile robot. For instance, we can notice that the functional *building_1* is comprised – among others – of functional *wall_1F*, whose architectural equivalent *wall_1* is externally connected to *wall_2* (both walls are disconnected from *asphalt_1*).

Concerning point 2), HORUS has been implemented in the SCONE KBS, serving as a knowledge repository and inference engine of the ACT-R cognitive model for building prediction: in particular, chunk-types for walls, building patterns, surrounding objects and urban features, have been designed on the basis of the conceptual specifications of HORUS and properly rendered into ACT-R declarative semantic chunks. The mechanism of building pattern recognition based on HORUS leverages a combination of general common-sense knowledge and specific topological features of given test-sites. In the medium term, we are planning to rely on diversified use cases from FTIG realistic scenario to implement and assess cognitive learning mechanism for recognizing previously unseen buildings, eventually storing new emergent patterns in HORUS as the result of a cognitively-based knowledge acquisition process.

Figure 2 – FTIG test-site.

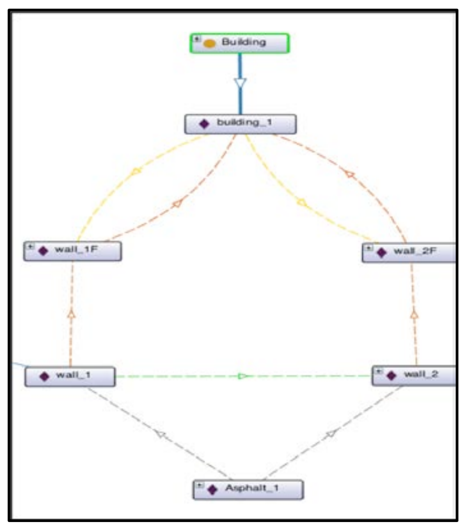

Figure 3 – *Example of distinction between features buildings.*

3.3 Functional Model of Building Prediction

The diagram in Figure 4 shows at the functional level how the cognitive system elaborates perceptual information. In particular, the human-like capability of recognizing complex objects– in our specific case buildings – on the basis of component parts and relative spatial properties is emulated by the pattern matching mechanisms realized by ACT-R cognitive architecture synchronized with SCONE KBS. In the real application scenario, the cognitive system looks up perceptual information in the World Model: for the sake of simplicity, we just assume here that semantically labeled perceptual inputs are directly fed into the cognitive system. The ACT-R model encodes perceptual inputs in declarative memory, associating them to pre-defined chunk-types, the most relevant ones being "Wall" and "Building". The chunk type "Wall" contains metric information represented by appropriate slots, e.g. height, width, length, and orientation. The chunk type "Building" denotes known configuration of the building via a set of specific walls, some invariant structures of the surroundings like a "Sidewalk", some "Gravel", a "Pole" or a "Dumpster"[4]. After the encoding phase, which in a real-world scenario is reiterated as long as the robot moves and perceives the environment, the actual procedures of recognition trigger: in particular, *partial matching* is used to compare input walls to the set of walls that constitute building patterns. This comparison is driven by metric characteristics: similarity is defined with respect to the difference between the dimensions of a perceived wall and the dimensions stored in declarative memory. The model is currently able to compute the difference of length and height: the smaller the difference, the more likely the perceived wall belongs to the building pattern containing a wall with the same dimensions. But in reality, wall-size is not always sufficient to discriminate which is the best matching building: for instance, hotels and hospitals may look akin in terms of shape and wall composition, despite having completely different purposes. In these regards, it may be beneficial for the recognition procedures to also process distinctive features of the building surroundings, like signs, benches, parking lots, heliports, sidewalks, gravel, ground etc. In the example illustrated in figure 4, few walls are detected by the robot vision systems, together with a car and a sidewalk. In order to enhance the overall recognition/prediction, the ACT-R model evokes the SCONE KBS through the dedicated module, querying HORUS about whether the collected perceptual contents instantiate any significant ontological relationship. In particular, tests are made to check for 1) *equivalence*, the fact that two entities are of the same kind (e.g. car, sedan) 2) *part-of*, the fact that an entity is part of another (wheel, car), and 3) *feature*, defined as a special "parasitic" part that exists in virtue of the host object (e.g. a bump of a road, a hole in a shoe, the bottom of the table, and – in our context – all the features of a building, e.g. a window in the wall, the grass porch of a house, the sidewalk in front of the entrance door, etc. – (Simons, 2000) and (van Inwagen, 1995). At the architecture level, these tests correspond to a sequence of comparable procedures, where only the contents of the SCONE buffer change according to the type of ontological query: in this sense, further tests can be easily added to the system as specific realizations of the same reusable production schema.

As depicted in Figure 4, the query trivially fails for the first two tests (NIL value returned), whereas the fact that the detected sidewalk is a feature of the detected wall W_1 is confirmed as being a true statement in the ontology HORUS – which means that it has been included as an

[4] These configurations are harvested from a database containing topological and morphological information about a given test-site.

assertion or 'A-Box' in the ontology (Sowa, 1984). In our model, a validation of perceptual cues by background knowledge prompts the ACT-R sub-symbolic mechanisms of *spreading of activation*, so that the related chunks (in this example only "sidewalk") are used to boost the activation of the "consistent" building pattern chunks, penalizing the inconsistent ones. ACT-R chunk activation is calculated by the following equation:

$$A_i = \ln \sum_j t_j^{-d} + \sum_k W_k S_{ki} + \sum_l MP_l Sim_{li} + N(0, \sigma)$$

On the basis of the first term, the more recently and frequently a chunk i has been retrieved, the higher the activation and the chances of being retrieved (t_j is the time elapsed since the j^{th} reference to chunk i and d represents the memory decay rate). In our scenario this would correspond to the priority accounted to more recent perceived walls over older visual cues. In the second term of the equation, the contextual spreading activation of a chunk i is set by the attentional weight W_k, given the element k and the strength of association S_{ki} between k and the i (the more "consistent" k and i are, the higher will S_{ki} be). The third term states that, under partial matching, ACT-R can retrieve the chunk that matches the retrieval constraints to the greatest degree, combining the similarity Sim_{li} between l and i (a negative score that is assigned to discriminate the 'distance' between two terms) with the scaling mismatch penalty MP (in our case similarity between walls is based on metric dimensions). The final factor adds a random component to the retrieval process by including Gaussian noise to make retrieval probabilistic.

We've previously used the quotation marks to highlight that the notion of "consistence" into play here goes beyond pure logical formalisms, involving cognitive plausibility weighted through stochastic mechanisms: this is a noteworthy aspect of using a non-algorithmic cognitive approach to problem solving, where reasonableness of results replaces optimality. The unified output of the reasoning procedures can be expressed by a probability distribution over the activated building patterns, where the most active pattern is "Hotel" in the example provided by Figure 4.

.

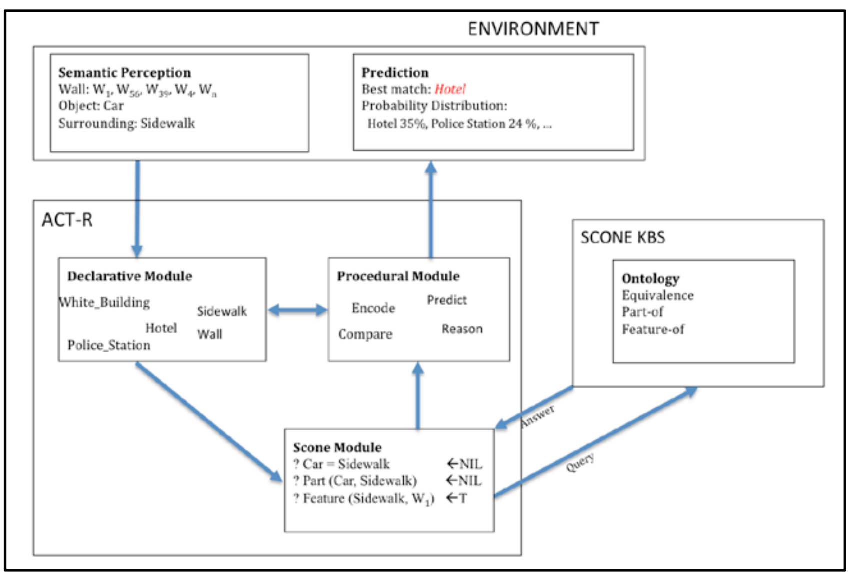

Figure 4 Diagram of The Cognitive Model for Building Prediction.

4. Evaluation

The current capabilities of the hybrid cognitive system have been analyzed in a synthetic simulation designed to cover alternate perceptual conditions: in this respect, the test that we conducted assumes that the semantic perception modules can classify objects in the field of view without ambiguity or degradedness[5].

The evaluation set consists of the 14 different buildings located in the FTIG test site, including "restaurant", "church", "fire station", "crypt", "ops storage", "police station", "mayor's mansion", etc.: despite the distinctive semantic labels, the *inherent differences* within buildings mainly concern the metric level, in particular the height and width of constituent walls. Consequently, we designed differential experimental settings with the purpose of gauging the effect of variable levels of complexity in the prediction process, hypothesizing that the robot progressively perceives one wall, two walls, three walls and four walls. This sequence of visual inputs denotes incremental levels of completeness with respect to predefined building patterns, although – as we'll see in the experiments – permutations of walls from different building patterns are also explored. In addition to metric similarities, we tested the extent to which, validating *extrinsic differences* within buildings by means of the HORUS ontology increases the probability of retrieving the correct pattern. In compliance with this premise, the ontology reflects the constraint that each building can be associated to no more than two relevant features, spanning from diverse types of terrain (e.g., "grass", "asphalt", "gravel" and "concrete"), to significant types of connected parts like "stairs", "doors", "windows" and characteristic contiguous objects, e.g. "gas pump" (which actually occurs only in the pattern associated to the "gas station"), "tree", "pole" or "car" (whose multiple instances are usually located nearby parking lots or garages). The association between buildings and features mirrors the actual spatial arrangement of related entities in the FTIG environment. In the first experiment (Figure 5), we fed the cognitive system with all the permutations of the walls constituting a single building (in the specific case "restaurant"), alternatively generating a consistent feature ("stairs") or an inconsistent one ("tree"). The knowledge component is triggered and combined with metric computations only when the consistent object is encoded and recognized as such by the hybrid cognitive system (using SCONE), whereas only metric computations activate when inconsistent objects are recognized as such. In this regard, we refer to the combined recognition mechanism as "contextual" prediction, whereas the metric-based similarity evaluation is referred to as "context-free" prediction. This distinction, which applies to all the experiments presented in this section, is exposed by figures 5-8, where 1) the graphs on the left columns represent the average results across visual conditions for the context-free predictions (where only *partial matching* based on walls dimensions is used), whereas 2) the graphs on the right column depend on factoring in contextual reasoning validation (which translates to *spreading of activation* from the detection of consistent objects to related building patterns in the prediction mechanism). From a cognitive point of view, in the experimental setting 2) the ontology is effectively used to determine the representation of the context, with elements that are judged as unrelated or incompatible excluded from the pattern recognition process, and thus unable to spread activation. In the second

[5] At the time of writing the authors are assembling the Intelligent Architecture for the annual assessment of the RCTA program, where the cognitive system will be actually deployed in a robotic infrastructure and experimentally evaluated in a test site.

experiment (Figure 6), we augmented the complexity of the test by generating all the permutations to walls of two different buildings, the “mayor’s mansion” and “service station”, which are relatively similar in terms of the height and width of walls but obviously different, from a human standpoint[6], with respect to semantic connotations and telicity[7]. In this case, we assessed that the recognition of the “gas pump” clearly helps to disambiguate the context, especially in the less informative situations, where only one or two walls are detected. This remark generally holds for every experimental condition that we considered: the higher the number of walls successfully associated to the correct building, the smaller the discriminatory effect of the ontology (though it can still be crucial for the refinement of specific algorithms – e.g. path planning). For the sake of testing increasing complexity, in the third experiment, wall permutations relative to three buildings were generated (“garage”, “townhouse” and “crypt”) and alternately coupled with two features (“opening” and “grass”), respectively consistent with the first two different buildings. As figure 7 indicates, while the process based on purely metric similarity was in average retrieving the “townhouse” as the most likely candidate, adding the ontology-driven cognitive mechanism allows discarding that result and boosting the recognition of the “garage” and the “crypt”. “Grass” was also a pertinent feature of the “church” building, and this is reflected by the probability of activation of the corresponding pattern (labeled as “C” at the extreme left of the figure). The final experiment concerned the “church”, the most complex building of the test set, whose peculiar configuration is generally unique in the FTIG site. Although being morphologically distinctive in a given environment might be seen as an unconditional advantage for the prediction task, in practice the unusual spatial configuration of the walls, the singular presence of a bell tower as part of the building and the heterogeneous heights and widths of the constituent walls imply that the cognitive system needs to analyze an extensive information stream, in terms of geometric characteristics, conceptual and temporal structures. This complexity is clearly reflected by the results reported in the left columns of figure 8, where the probability value of the church pattern is slightly higher than the others, but overall extremely low in all four conditions. The right column shows, instead, that validation by knowledge boosts the value of the predicted patterns, even if - by using the shared surrounding feature “concrete” - the two patterns for “crypt” and “restaurant” are almost equally retrieved. Because of the computational requirements of this knowledge-intensive simulation framework, we could not extend the experimentation beyond wall permutations of three buildings and three objects, although we aim at reporting these results in future work. Nevertheless, we think that the general trends of the predictions properly reflect the benefits of using ontology-based cognitive reasoning in conjunction with probabilistic computations, notwithstanding the intrinsic limits of assessing the cognitive system independently from the deployment in robots.

[6] Since the final objective of the cognitive system is to be deployed in a robot working with soldiers as a team member, sharing mental models with them is important.

[7] Telicity denotes the property of operational/social functionality assigned to artifacts.

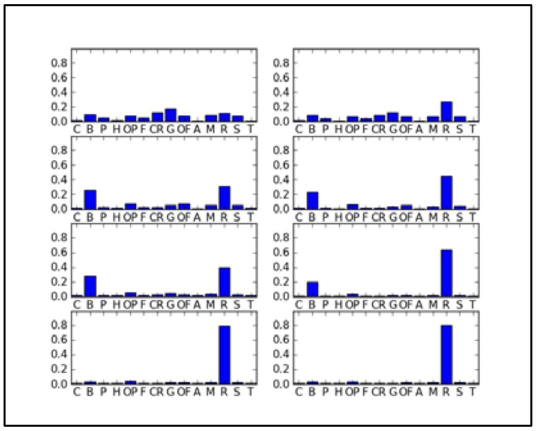

Figure 5

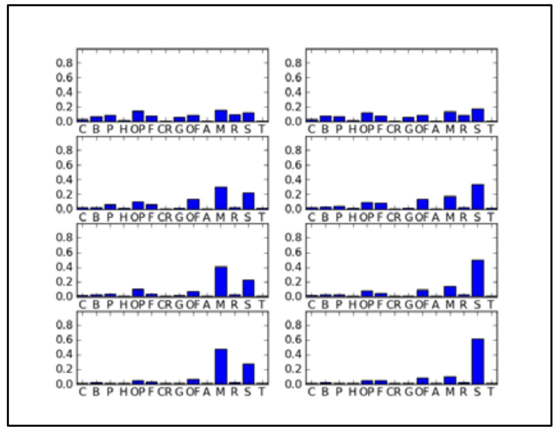

Figure 6

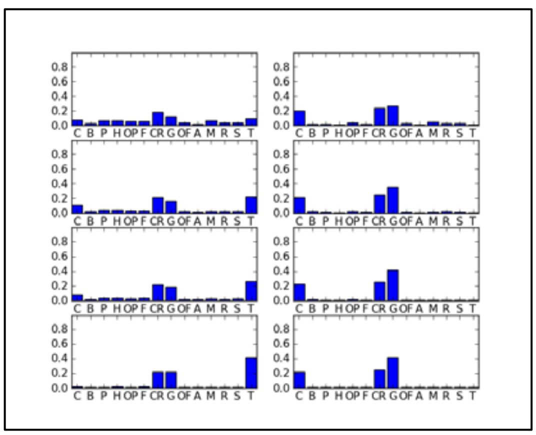

Figure 7

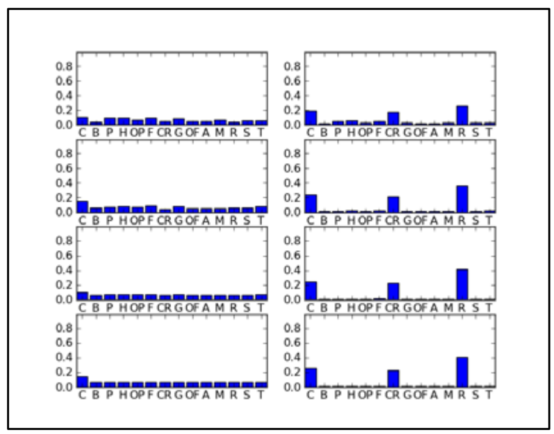

Figure 8

*LEGENDA**:** Y-axis: probability associated to building predictions; X-axis: labels for the14 buildings considered (C: Church; B: Bar; P: Police Station; H: Hotel; OP: Ops Storage; F: Fire Station; CR: Crypt; G: Garage; OF: Office Building; A: Aid Station; M: Mayor's House; R: Restaurant; S: Service Station; T: Townhouse.*

5. Conclusion

In this paper we presented a hybrid cognitive system for high-level recognition of buildings in a robotic framework. We showed how the integration between cognitive architectures and ontologies can leverage human-like capabilities in artificial agents. Future work will be devoted to a broad validation of the system with respect to a dataset of real-world scenarios, where the hybrid system will be enhanced both in terms of learning capabilities and quantitative spatial reasoning.

Acknowledgements

The authors thank Martial Hebert, Arne Suppe and Unmesh Kurup for providing inspired ideas and insightful discussions. This work was conducted through collaborative participation in the Robotics Consortium sponsored by the U.S Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement W911NF-10-2-0016.

References

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.

Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought.* Erlbaum.

Bateman, J., & Farrar, S. (2005). Modelling models of robot naviagtion using formal spatial ontology. In C. Freksa, M. and Knauff, B. Krieg-Brückner, B. Nebel, & T. Barkowsky (Ed.), *Spatial Cognition IV: Reasoning, Action, Interaction. International Conference Spatial Cognition.*, (pp. 366-389). Berlin.

Biederman, I. (1987). Recognition-by-components: a Theory of Human Understanding. *Psychological Review* , (94), 115-147.

Borgo, S., & Masolo, C. (2009). Ontological Foundations of DOLCE. In S. Staab, & R. Studer, *Handbook on Ontologies - 2nd Edition.* Berlin: Springer Verlag.

BouSaba, C., Esterline, A., Homaifar, A., & Fatehi, F. (2008). Spatial Ontologies for Tactical Behaviors. In G. Gerhart, & D. S. Gage (Ed.), *Unmanned Systems Technology X.* Orlando (FL).

Casati, R., & Varzi, A. (1996). *Events* . Darthmouth, USA.

Cassimatis, N., Trafton, J., Bugajska, M., & Schultz, A. C. (2004). Integrating cognition, perception and action through mental simulation in robots. *Robotics and Autonomous Systems* (49), 13-23.

Fahlman, S. (2006). Using SCONE's multiple-context mechanism to emulate human-like reasoning. *First International Conference on Knowledge Science, Engineering and Management (KSEM).* Springer-Verlag.

Guarino, N. (1998). Formal Ontology and Information Systems. *Formal Ontology in Information System.* IOS Press.

Hawes, N., & Klenk, M. L. (2012). Towards a Cognitive System That Can recognize Spatial Regions Based on Context. *26th National Conference on Artificial Intelligence* .

Kelly, T., & Avery, E. (2010). A cognitive robotics system: the symbolic and sub-symbolic robotic intelligence control system (SS-RICS). *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications (SPIE)*, (pp. 110-140). Orlando (FL).

Kurup, U., & Lebiere, C. (2012). What can cognitive architectures do for robotics? *Biologically Inspired Cognitive Architectures , 2*, 88-99.
Kurzweil, R. (2012). *How to Create a Mind.* Viking.
Kutz, O., Lücke, D., & Mossakowski, T. .. (2008). Heterogeneously Structured Ontologies—Integration, Connection, and Refinement. In T. Meyer, & M. Orgun (Ed.), *Knowledge Representation Ontology. Workshop* (pp. 41-50). ACS.
Moravec, H. (1988). *Mind Children.* Cambridge, Massachusetts, USA: Harvard University Press.
Simon, H. (1991). Bounded Rationality and Organizational Learning. *Organization Science , 2* (1), 125-134.
Simons, P. (2000). *Parts: A Study in Ontology.* Oxford, UK: Oxford University Press.
Sowa, J. (1984). *Conceptual Structures: Information Processing in Mind and Machine.* Reading: Addison Wesley.
Staab, S., & Studer, R. (2004). *Handbook on Ontologies.* Springer.
Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional Routing of Information to the Cortex: A Model of the Basal Ganglia's Role in Cognitive Coordination. *Psychological Review , 117* (2), 541-574.
Trafton, J., Hiatt, L., Harrison, A. M., & Khemlani, S. S. (2012). ACT-/E: An embodied cognitive architecture for Human-Robot Interaction. *Journal of Human-Robot Interaction , 1* (1), 78-95.
Tversky, A. (1977). Features of Similarity. *Psychological Review* (84), 327-352.
van Inwagen, P. (1995). *Material Beings* . Ithaca and London.: Cornell University Press.