Towards a Cognitive System that Can Recognize Spatial Regions Based on Context

Nick Hawes
Intelligent Robotics Lab
University of Birmingham, UK
n.a.hawes@cs.bham.ac.uk

Matthew Klenk
Palo Alto Research Center
Palo Alto, CA
matthew.klenk@parc.com

Kate Lockwood
ITCD Department
California State University - Monterey Bay
klockwood@csumb.edu

Graham S. Horn
Intelligent Robotics Lab
University of Birmingham, UK
gsh148@cs.bham.ac.uk

John D. Kelleher
Applied Intelligence Research Centre
Dublin Institute of Technology
john.d.kelleher@dit.ie

Abstract

In order to collaborate with people in the real world, cognitive systems must be able to represent and reason about spatial regions in human environments. Consider the command *"go to the front of the classroom"*. The spatial region mentioned (the front of the classroom) is not perceivable using geometry alone. Instead it is defined by its functional use, implied by nearby objects and their configuration. In this paper, we define such areas as *context-dependent spatial regions* and present a cognitive system able to learn them by combining qualitative spatial representations, semantic labels, and analogy. The system is capable of generating a collection of qualitative spatial representations describing the configuration of the entities it perceives in the world. It can then be taught context-dependent spatial regions using *anchor points* defined on these representations. From this we then demonstrate how an existing computational model of analogy can be used to detect context-dependent spatial regions in previously unseen rooms. To evaluate this process we compare detected regions to annotations made on maps of real rooms by human volunteers.

Introduction

Consider a janitorial robot cleaning a classroom. While performing this task, it encounters a teacher working with a student. The teacher tells the robot to "start at the front of the classroom", expecting it to go to the front of the classroom and begin cleaning that area. This response requires that the robot is able to *determine the spatial region in the environment that satisfies this concept*.

The ability to understand and reason about *spatial regions* is essential for cognitive systems performing tasks for humans in everyday environments. Some regions, such as whole rooms and corridors, are defined by clearly perceivable boundaries (e.g. walls and doors). However, many regions to which humans routinely refer are not so easily defined. Consider, for example, the aforementioned region *the front of the classroom*. This region is not perceivable using just the geometry of the environment. Instead, it is defined by the objects present in the room (chairs, a desk, a whiteboard), their role in this context (seats for students to watch a teacher who writes on the whiteboard) and their configuration in space (the seats point toward the whiteboard). We refer to such regions as *context-dependent spatial regions* (CDSRs).

Current cognitive systems are not capable of representing and reasoning about CDSRs, yet it is an important ability. If cognitive systems are to collaborate with humans in everyday environments then they must be able to understand and refer to the same spatial regions humans do. Many regions are best defined in a context-dependent manner, for example, a kitchen in a studio apartment, an aisle in a church or store, behind enemy lines in a military engagement, etc. In order to represent and reason about such regions, cognitive systems must integrate different types of information, including geometric, semantic, and functional knowledge. Creating systems able to integrate such a range of information is a key challenge in the cognitive systems paradigm (Langley in press).

This paper presents an artificial cognitive system (specifically a mobile robot) able to represent and reason about CDSRs. Our approach is founded on two assumptions. The first assumption is that CDSRs can be defined using *qualitative spatial representations* (QSRs) corresponding to sensor data of the system (Cohn and Hazarika 2001). The second assumption is that semantically and geometrically similar areas (e.g. two different classrooms) will feature similar CDSRs, and that these similarities can be recognised through *analogy*. The rest of the paper is structured following these assumptions. Section describes how we generate QSRs from sensor data taken from an existing, state-of-the-art, cognitive system and use these to define CDSRs. Section then describes how we use the structure-mapping model of analogy (Gentner 1983) to transfer a CDSR from a labelled example to a new situation. Section presents a worked example of the entire process, and Section evaluates our system in comparison to data from human subjects performing the same task.

Metric to Qualitative Representations

The context which defines a CDSR is a combination of the functional and geometric properties of a room, i.e. what can be done there and where. In this work we implicitly represent context using the types of objects present in a room and their location relative to each other. The following sections describe how we construct symbolic representations of these ingredients of context from robot sensor data.

The Dora System

We base our work on Dora, a mobile cognitive robot with a pre-existing multi-layered spatial model (Hawes et al. 2011). In this paper, we draw on the metric map from this model. For more information on Dora's other competences, see recent papers, e.g. (Hawes et al. 2011; Hanheide et al. 2011).

Dora's metric map is a collection of lines in a 2D global coordinate frame. Two example maps are pictured in Figure 4. Map lines are generated by a process which uses input from the robot's odometry and laser scanner to perform simultaneous localization and mapping (SLAM). Lines in this SLAM map represent features extracted from laser scans wherever a straight line is present for long enough to be considered permanent. In practice, lines are generated at the positions of walls and any other objects that are flat at the height of the laser (e.g. bins, closed doors etc.). The robot's location in the metric layer is represented as a 2D position plus an orientation.

Dora is capable of using vision to recognize pre-trained 3D object models. Recognition can either be triggered through autonomous visual search or at a user's command. When an object is detected it is represented in the metric map by placing a copy of the model at the detected pose. The recognizer associates each object with a semantic type that was provided during a training phase.

To enable us to generate a range of different evaluation situations in a reasonable length of time, we have generated data from Dora in both real rooms and in simulation. Simulation is performed using the Player/Stage hardware abstraction layer (Gerkey, Vaughan, and Howard 2003) allowing us to run the system mostly unchanged in a predefined environment. Also, to enable us to detect a wider range of objects than is usually possible (from armchairs to whiteboards), we used a simulated object recogniser in all runs. The recogniser was configured with types and positions of objects in the environment and was guaranteed to detect them when the robot was orientated towards them. This eliminated any errors from the recognition process, but was still influenced by errors in robot localisation.

Qualitative Spatial Representation Extraction

For each object that Dora detects we compute the strengths of 8 spatial relations between that object and each of the objects adjacent to it; adjacency is determined using a voronoi diagram, as is standard in geometric reasoning (Forbus, Usher, and Chapman 2003). The strength of a computed relation for a given pair of objects represents the applicability of that relation to the pair. Strength ranges from 0 to 1, with 0 being unsuitable. The model used to compute these relations was inspired by the literature on modeling the semantics of spatial terms (Kelleher and Costello 2009; Kelleher and van Genabith 2006; Regier and Carlson 2001; Gapp 1994). The model accommodates both direction and distance as factors in the relative position of objects.

The relations we compute between each given *landmark* object and its adjacent neighbours are analogous to the cardinal and intermediate points on the compass when the compass is centered on the object. The canonical directions of these relations are defined using the following vectors: $\langle 0,1\rangle$, $\langle 1,1\rangle$, $\langle 1,0\rangle$, $\langle 1,-1\rangle$, $\langle 0,-1\rangle$, $\langle -1,-1\rangle$, $\langle -1,0\rangle$, $\langle -1,1\rangle$. The predicates used to denote these relations are named accordingly, e.g. *xZeroYPlus*, *xPlusYPlus*, *xPlusYZero*, *xPlusYMinus*, etc.

We generate the strengths of these spatial relations as follows. First we compute the maximum distance d_{max} between any two points in the room, this value is used to normalize the distances between objects. Next, taking each object in turn to be the landmark, we translate the origin of the room to the landmark's centroid. This results in the coordinates of the all the other objects in the room being translated into a frame of reference whose origin is the centroid of the landmark. We then compute the strength of each of the 8 spatial relations between the landmark and each of the objects adjacent to it by calculating: (a) the distance d between the landmark's centroid and the adjacent object's location, and (b) the inner angle θ between the direction vector of the relation and the vector from the origin (the landmark's centroid) to the neighbour's location. These two spatial components are integrated to compute the strength s of a given relationship using Equation 1. Figure 1 provides a visualization of a spatial relationship across a region.

$$s = \begin{cases} \left(1 - \frac{\theta}{90}\right) * \left(1 - \frac{d}{d_{max}}\right) & \text{if} \quad \theta \leq 90^\circ \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

These spatial relationships between adjacent objects provide the structure necessary for analogical processing. Generating the relationships in this way (as opposed to, for example, simple coordinate-based thresholding) has the advantage that the presence and absence of relationships is represented on a continuous scale. This provides our representations with the flexibility necessary to manage the variation in perceptual information (i.e. the position of walls and objects) inevitable in human environments and robot perception.

In addition to spatial relations, we also create *grouping entities* from the robot sensor data. Grouping entities collect together sets of adjacent objects of the same type. For example, a classroom would likely have a group entity created in which all of the students' desks were members.

Representing CDSRs

We use *anchor points* (Klenk et al. 2005) to define the boundaries of CDSRs. Anchor points are symbolic descriptions which link a conceptual entity to a perceived entity. The perceived entities we use are the objects recognised by Dora, and the room itself. The room representation is

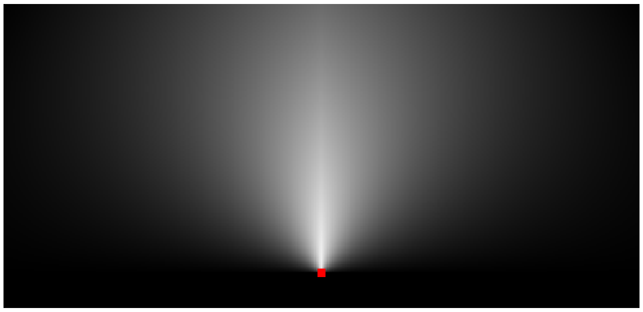

Figure 1: A visualisation of a the strength of a spatial relation across a region. The landmark is the red square and the direction vector used was $\langle 0, 1 \rangle$ (i.e. above of the landmark). The lighter the pixel the stronger the spatial relation is deemed to be at that point.

created by putting a convex hull around the lines in Dora's SLAM map. Anchor points are created from perceived entities using unary functions, e.g. `(XMaxYMostFn Desk1)` represents the point on the `Desk1` with the largest x coordinate taken from the set of points with a y coordinate within 5% of the maximum y coordinate. Anchor points are linked to particular CDSRs using a `boundarySegment` ternary relation. After we have defined the boundary of the region, we assign it a semantic label using the `regionType` relation. Therefore, each CDSR has one type and a variable number of boundary segments.

```
(regionType CDSR9 FrontRegion)
(boundarySegment CDSR9
    (YMaxXFewestFn Room3)
    (YMinXFewestFn Room3))
(boundarySegment CDSR9
    (YMinXFewestFn Room3)
    (YMinXFewestFn Group1))
```

Figure 2: Three of the five expressions representing the front of the classroom context-dependent region `CDSR9`

Figure 2 contains three of the five expressions defining the front of classroom `Room3` which is pictured in the top of Figure 4. The boundary segments (shown in orange in Figure 4) define the extent of the region. `(YMaxXFewestFn Room3)` and `(YMinXFewestFn Room3)` are the points with the highest and lowest y coordinate out of the set of points within 5% of the minimum x coordinate of `Room3`. The next segment connects the lower left coordinate in the figure to the `(YMinXFewestFn Group1)`, where `Group1` includes the eight desks. There are two more boundary segments completing a polygon for this region. The semantic label `FrontRegion` ties this polygon to a conceptual region, "the front of the room". This definition for the front of the room is specific to `Room3` and its entities. It is clearly context-dependent because its extent is dependent on the arrangement of the anchor points used to define its boundary. If the desks were in a different position then the region would cover a different extent (e.g. if they were further to the left then the region would be smaller).

Analogical Transfer of Spatial Regions

We assume that a cognitive system will have a way of initially acquiring examples of CDSRs, e.g., by being taught through dialogue, sketching, or hand-coding. To avoid burdening potential users with the task of teaching the system every CDSR individually, it is desirable for a cognitive system to be able to automatically recognize similar regions after initial training. For example, after a janitorial robot has been taught where the front of one classroom is, it should be able to identify the fronts of other classrooms in the building. Our system uses *analogy* to solve this problem. We chose this approach because analogy has been previously used to successfully combine semantic and geometric information in spatial reasoning tasks (Lockwood, Lovett, and Forbus 2008).

Analogy is an essential cognitive process. In humans, analogical processing has been observed in language comprehension, problem-solving, and generalization (Gentner 2003). The structure-mapping theory of analogy and similarity postulates this process as an alignment between two structured representations, a *base* and a *target* (Gentner 1983). We use the Structure-Mapping Engine (SME) (Falkenhainer, Forbus, and Gentner 1989) to perform analogical matching in our system. Given base and target representations as input, SME produces one or more mappings. Each mapping is represented by a set of *correspondences* between entities and expressions in the base and target structures. Mappings are defined by expressions with an identical relation and corresponding arguments. When provided with expression strengths, such as, our spatial relationships, SME prefers mappings with closely aligned fact strengths. SME can be given *pragmatic constraints* that require certain entities in the base to be included in the mapping. Mappings also include *candidate inferences* which are conjectures about the target using expressions from the base which, while unmapped in their entirety, have subcomponents that participate in the mapping's correspondences. SME operates in polynomial time, using a greedy algorithm (Forbus, Ferguson, and Gentner 1994).

Figure 3 illustrates a sample mapping between six base expressions and three target ones. Each oval represents a predicate, and the entity arguments are represented by squares. SME generates a mapping between the base expressions `(group Desk1 Desk2)` and `(xMinusYZero Desk1 Desk2)`, and the target expressions `(group Desk11 Desk12)` and `(xMinusYZero Desk11 Desk12)` as well as between the `regionType` expressions in each case in the following manner. First, the predicates of these expressions are placed in correspondence, as identical predicates are preferred by SME. Then SME aligns the arguments of the aligned predicates, `Desk1` with `Desk11`, `Desk2` with `Desk12`, and `CDSR1` with `CDSR2`. While there is another `XMinusYZero` statement in the base about two desks, it cannot correspond to either of the target expressions in the same mapping due to the one-to-one constraint in SME which allows each element in the target to map to at most one element in the base and vice

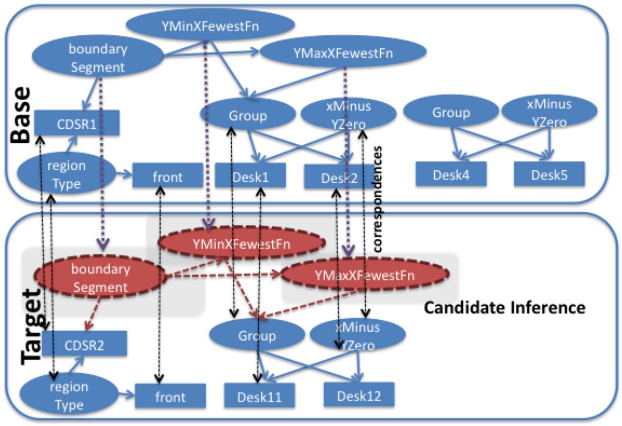

Figure 3: Analogical mapping between six base expressions and three target expressions.

versa. In Figure 3, the correspondences are highlighted by the hashed bi-directional arrows. Next, SME creates a candidate inference for the boundary segment expression, because both the mapped `Group` and `regionType` predicates participate in the mapping. The candidate inference is shown in red in the figure. Note that inference is selective, with no candidate inferences generated for the entirely unmapped expressions.

In our system, the base and target representations consist of the entities Dora has perceived in two different rooms, the QSRs between them and any groups that have been identified. The base also contains a labeled CDSR of the type sought in target. The result of running SME on these representations is a set of correspondences between the base and target, and a set of candidate inferences about the target. We use these to transfer the CDSR from base to target (i.e. recognizing the CDSR in the target) as follows. First, we identify the CDSR of the sought type in the base and use SME's pragmatic constraints to ensure that the entities referred to its anchor points participate in the mapping. To transfer the CDSR to the target, we collect the candidate inferences that result from `boundarySegment` statements mentioning the base CDSR. The second and third arguments of these candidate inferences are anchor points in the target environment. We use these to define the boundary of the CDSR in the target.

Example System Run

To elucidate the workings of our system, we now present an example of how it can transfer a CDSR describing the front of a known classroom (the base) to a new classroom (the target).

We first create the base and target representations by running Dora in the two different classrooms. In each case, Dora is manually driven around the room to allow it to create a metric map. Once the map is created, Dora is then positioned such that the objects are observable and the visual recognition system is run. The map and object data that result from this are then passed on to the QSR generator.

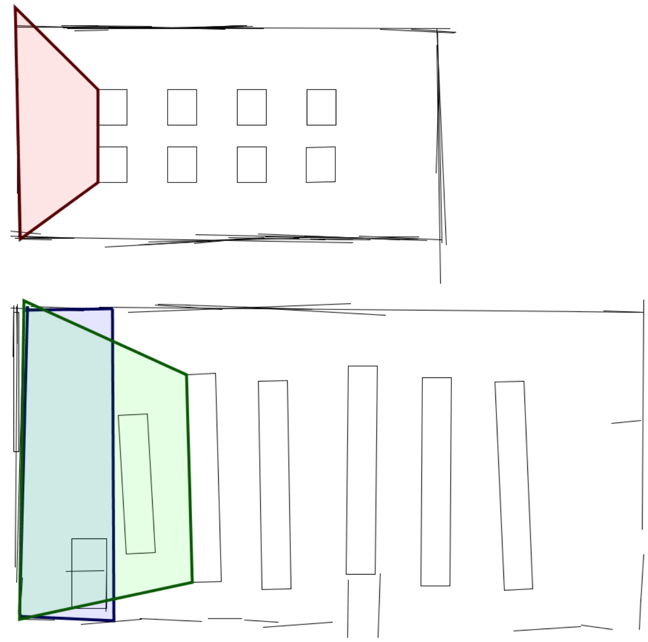

Figure 4: Maps of 2 real classrooms generated by our system. The lines around the perimeter are walls, the unfilled polygons are the outlines of objects and the filled polygons are CDSRs. The maps show an expert-annotated CDSR (red, top image), a subject-annotated CDSR (blue, bottom image) and a CDSR transferred by analogy (green bottom image). The classroom used to generate the bottom classroom is pictured in Figure 5.

The base and target maps are pictured in the top and bottom of Figure 4 respectively. In the base case, Dora perceives 8 individual desks, a group entity containing these desks and the room area. To this we add the CDSR representing the front of the room. The case includes a total of 50 expression relating the 20 entities. Six of these expressions are used to define the boundary segments and CDSR representing the front of the room. The target case includes 26 expressions and 11 entities.

SME generates an analogy between the base and target cases enabling the transfer of the symbolic description of the front of the room to the new situation requiring `Room3` and `Group1` participate in the mapping as they are referenced by the anchor points in the base. The resulting analogy includes 26 correspondences between the entities and expressions and 32 candidate inferences. Four of these candidate inferences define the CDSR in the target with anchor points defined on the room and the group of desks in the target. The green region in the lower image of Figure 4 illustrates the transferred CDSR.

Evaluation

To evaluate our progress toward building a cognitive system capable of reasoning about CDSRs, we conducted the an experiment focusing on the following questions:

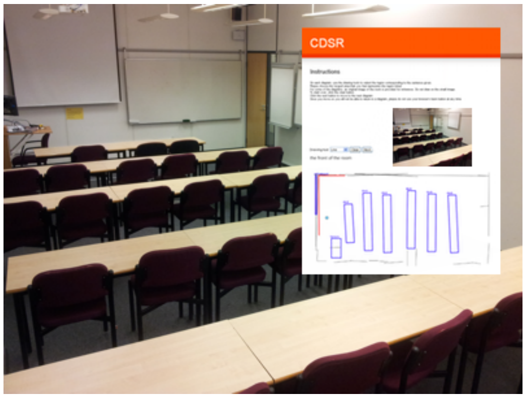

Figure 5: One of the classrooms used in our evaluation. This image was presented to subjects who were asked to annotate a copy of the image in the bottom half of Figure 4. The inset shows a screenshot from the data collection webpage.

- Are anchor points able to encode context-dependent spatial regions?
- When provided with a base representation containing a labelled CDSR, how well does our approach identify the CDSR in a given target?

Materials

We evaluated our approach on six classrooms (two simulated and four real) and two simulated studio apartments. The simulated rooms were based on real-life counterparts. For each room we manually encoded appropriate CDSRs that could be represented by our approach. For the classrooms these were the front and back, and the front and back rows of desks. For the studios these were the kitchen, office and living areas. These manually encoded regions were used as the base CDSRs for analogical transfers, and can be considered the training data for our evaluation.

To determine how people define CDSRs, we asked three naïve users to draw polygons for each region type for each room. This task was performed using a webpage on which each user was presented with an image of the real room plus an image of the map data generated by the robot onto which the drawing could be done. The webpage[1] is shown in the inset in Figure 5. The user-defined polygons define the *target regions* against which we evaluate our transfers.

We consider a *problem instance* to be a room and a sought CDSR type. For each room containing a manually encoded CDSR of the sought type, we generate a *transferred region* using analogical transfer. To assess the quality of the transfer, we calculate precision (p, the proportion of the transferred region that overlaps with the target region) and recall (r, the proportion of the target region that overlaps with the transferred region) as follows:

$$p = \frac{area(transferred\ region \cap target\ region)}{area(transferred\ region)} \qquad (2)$$

[1]http://home.csumb.edu/k/katherinelockwood/world/

$$r = \frac{area(transferred\ region \cap target\ region)}{area(target\ region)} \qquad (3)$$

Using this approach we generate results showing the matches between each of the following pairs of regions: the transferred region and the appropriate target region; the CDSR we manually encoded for the target room and target region; and the region for the whole room and the target region. Results comparing transferred and target regions measure how well our system is able apply a single example to new situations. The comparisons between the manual annotations to the target regions measure how well the anchor points we chose capture the users' regions (who were not constrained to anchor points). Results from the whole room regions provide a baseline performance for comparison.

Results

To assess overall performance, Table 1 summarizes the results across all problem instances against user-defined target regions from three different users. The transferred regions achieved a precision of .47 (σ=.37) and a recall of .46 (σ=.38). Comparing the manually encoded regions against each target region results in a mean precision of .71 (σ=.30) and recall of .67 (σ=.25). The region defined by the room corresponds to the target region with a precision of .17 (σ=.11) and recall of .98 (σ=.05).

To identify how our approach fairs under different conditions, Table 2 separates the results by CDSR type. The mean precision for the transferred regions ranged from .76 for the front rows of classrooms to 0 for the office in studio apartments. Comparing manually encoded against target regions resulted in a minimum mean precision of .60. This occurred for the front of the classroom. The whole room precision, which is directly proportionally to the size of the target region, varied from .08 for the office to .35 for the living area.

Discussion

These results support the hypothesis that anchor points can provide a symbolic representation on top of sensor data for context-dependent spatial regions, and, when combined with qualitative spatial relations, they facilitate learning from a single example through analogical transfer. Collaboration with human users requires a high precision and recall, because cognitive systems must be able to understand as well as refer to these regions in human environments. Consequently, the high manually encoded precisions and recalls indicate that the defined anchor points are a reasonable starting point for a symbolic representation. Our future work seeks to further evaluate the utility of this representation by embedding the cognitive system within tasks with human users.

The transferred regions were considerably more precise (.47) when compared to the room as whole (.17), and their recalls (.46) indicate that they captured almost half of the area indicated by the human user. As we create CDSRs using anchor points defined on perceived entities, our approach performs best when the boundary of the target CDSR is closely tied to such entities. This is the case in the front rows of the classroom, with p of .76 and .82 for the inferred

Transferred	Manually Encoded	Entire Room
$\bar{p}$=.47 σ=.37, $\bar{r}$=.46 σ=.38	$\bar{p}$=.71 σ=.30, $\bar{r}$=.67 σ=.25	$\bar{p}$=.17 σ=.11, $\bar{r}$=.98 σ=.05

Table 1: Overall Performance Compared Against Target Regions Defined by Three Users

Region	Transferred	Manually Encoded	Entire Room
Front	$\bar{p}$=.32 σ=.33, $\bar{r}$=.49 σ=.41	$\bar{p}$=.60 σ=.29, $\bar{r}$=.83 σ=.19	$\bar{p}$=.16 σ=.10, $\bar{r}$=1 σ=0
Back	$\bar{p}$=.44 σ=.37, $\bar{r}$=.56 σ=.41	$\bar{p}$=.66 σ=.25, $\bar{r}$=.84 σ=.17	$\bar{p}$=.11 σ=.06, $\bar{r}$=.99 σ=.03
Front Rows	$\bar{p}$=.76 σ=.27, $\bar{r}$=.28 σ=.21	$\bar{p}$=.83 σ=.31, $\bar{r}$=.50 σ=.11	$\bar{p}$=.22 σ=.08, $\bar{r}$=1 σ=0
Back Rows	$\bar{p}$=.72 σ=.30, $\bar{r}$=.42 σ=.26	$\bar{p}$=.80 σ=.29, $\bar{r}$=.43 σ=.26	$\bar{p}$=.19 σ=.06, $\bar{r}$=1 σ=0
Kitchen	$\bar{p}$=.60 σ=.05, $\bar{r}$=.59 σ=.34	$\bar{p}$=.78 σ=.20, $\bar{r}$=.71 σ=.13	$\bar{p}$=.16 σ=.02, $\bar{r}$=.92 σ=.13
Office	$\bar{p}$=.00 σ=.00, $\bar{r}$=.00 σ=.00	$\bar{p}$=.78 σ=.29, $\bar{r}$=.55 σ=.20	$\bar{p}$=.08 σ=.03, $\bar{r}$=.94 σ=.06
Living Room	$\bar{p}$=.40 σ=.39, $\bar{r}$=.01 σ=.01	$\bar{p}$=.63 σ=.34, $\bar{r}$=.54 σ=.13	$\bar{p}$=.35 σ=.22, $\bar{r}$=.96 σ=.06

Table 2: Performance by Region Type

and the manually encoded regions respectively. The system performs worst when the extent of the CDSR is defined as an unbounded area near or around particular objects. The office of a studio apartment is loosely defined as the region around the desk. This motivates one direction of future work: expanding the vocabulary of anchor points to better capture these notions of space.

Related Work

Typical approaches to spatial representation for mobile robots tend to focus on localization, and thus mostly represent the world uniformly without subdivision into meaningful (semantic) units (Thrun 2003). When a more structured representation is required, many turn to Kuipers' Spatial Semantic Hierarchy (Kuipers 2000). This paper follows in this tradition, adding CDSRs to his qualitative topological representations. Whilst mobile robots exist which can determine the type of a room from the objects in it (Hanheide et al. 2010; Galindo et al. 2005), they only concern themselves with types of whole rooms, and cannot represent regions within rooms. This is also true for those systems which use some elements of QSR (Aydemir et al. 2011). The need for an autonomous system to ground references to human-generated descriptions of space has been recognized in domains where a robot must be instructed to perform a particular task, however existing systems are restricted to purely geometrically-defined regions (Tellex et al. 2011; Dzifcak et al. 2009; Brenner et al. 2007).

There is mounting evidence that analogy, operating over structured qualitative representations, can be used to simulate a number of spatial reasoning tasks. Forbus *et al.* showed that analogy between course of action diagrams could be used to identify potential ambush locations in new situations by focusing on only the relevant aspects of sketched battle plans (Forbus, Usher, and Chapman 2003). A core contribution of their work was the definition of a *shared similarity constraint* between a spatial reasoning system and its user; where users and spatial reasoning systems agree on the similarities between situations. This has close parallels to what we are trying to accomplish, where a cognitive system is able to reason about context-dependent spatial regions by identifying the same salient features as its human user. The anchor points in our work were originally used in teaching a system how to solve problems from the Bennett Mechanical Comprehension Test that require spatial and conceptual reasoning. For example, identifying which wheelbarrow will be more difficult to lift based on the relative locations of its loads as depicted in a sketch (Klenk et al. 2005). In that work, the anchor points defined the endpoints of lines. We go beyond that result to use anchor points to specify 2D regions.

Conclusion

In this paper we presented an integrated cognitive system capable of representing and reasoning about context-dependent spatial regions. The system identifies CDSRs in previously unseen environments through analogy with a single example. This is a difficult cognitive systems task requiring integration of semantic and geometric knowledge to identify regions as small as 8% of the room. Our system demonstrates a successful integration of a range of technologies including vision, SLAM, qualitative spatial reasoning and analogy to achieve this task. In order to make this rich collection of components work together, our work takes a number of short-cuts that we plan to address with future work. These include a reliance on the initial orientation of a room in a global coordinate frame, the lack of a mechanism to retrieve relevant rooms from memory (e.g. MAC/FAC (Forbus, Gentner, and Law 1995)), and a lack of transfer post-processing (e.g. comparing the QSRs present in both base and transferred regions) to improve results. In addition, we must complement our system development work with more comprehensive human studies assessing how people define and use these regions as well as how well anchor points capture them. Despite the preliminary nature of this work, our evaluation demonstrates that the system is able to transfer CDSRs that overlap with user-defined regions for 6 out of 7 region types.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement No.

215181, CogX. The authors would like to thank Jeffery Usher for providing the code to compute voronoi diagrams, Scott Friedman for assistance on probabilistic SME, and Danny Bobrow comments on an earlier draft of this paper.

References

Aydemir, A.; Sjöö, K.; Folkesson, J.; Pronobis, A.; and Jensfelt, P. 2011. Search in the real world: Active visual object search based on spatial relations. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA'11)*.

Brenner, M.; Hawes, N.; Kelleher, J.; and Wyatt, J. 2007. Mediating between qualitative and quantitative representations for task-orientated human-robot interaction. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI'07)*, 2072–2077.

Cohn, A. G., and Hazarika, S. M. 2001. Qualitative spatial representation and reasoning: an overview. *Fundam. Inf.* 46(1-2):1–29.

Dzifcak, J.; Scheutz, M.; Baral, C.; and Schermerhorn, P. 2009. What to do and how to do it: Translating natural language directives into temporal and dynamic logic representation for goal management and action execution. In *Proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA'09)*.

Falkenhainer, B.; Forbus, K. D.; and Gentner, D. 1989. The structure-mapping engine: Algorithm and examples. *Artificial Intelligence* 41(1):1 – 63.

Forbus, K. D.; Ferguson, R. W.; and Gentner, D. 1994. Incremental structure-mapping. In Ram, A., and Eiselt, K., eds., *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, 313–318.

Forbus, K.; Gentner, D.; and Law, K. 1995. MAC/FAC: A model of similarity-based retrieval. *Cognitive Science* 19(2):141 – 205.

Forbus, K.; Usher, J.; and Chapman, V. 2003. Qualitative spatial reasoning about sketch map. In *Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence*.

Galindo, C.; Saffiotti, A.; Coradeschi, S.; Buschka, P.; Fernandez-Madrigal, J. A.; and Gonzalez, J. 2005. Multi-hierarchical semantic maps for mobile robotics. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05)*, 2278 – 2283.

Gapp, K. P. 1994. Basic meanings of spatial relations: Computation and evaluation in 3d space. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI'94)*, 1393–1398. AAAI Press.

Gentner, D. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7(2):155 – 170.

Gentner, D. 2003. Why we're so smart. In Gentner, D., and Goldin-Meadow, S., eds., *Language in mind: Advances in the study of language and thought*. MIT Press. 195–235.

Gerkey, B. P.; Vaughan, R. T.; and Howard, A. 2003. The Player/Stage project: Tools for multi-robot and distributed sensor systems. In *Proceedings of the International Conference on Advanced Robotics (ICAR'03)*, 317–323.

Hanheide, M.; Hawes, N.; Wyatt, J.; Göbelbecker, M.; Brenner, M.; Sjöö, K.; Aydemir, A.; Jensfelt, P.; Zender, H.; and Kruijff, G.-J. M. 2010. A framework for goal generation and management. In *Proceedings of the AAAI'10 Workshop on Goal-Directed Autonomy*.

Hanheide, M.; Gretton, C.; Dearden, R.; Hawes, N.; Wyatt, J.; Pronobis, A.; Aydemir, A.; Göbelbecker, M.; and Zender, H. 2011. Exploiting probabilistic knowledge under uncertain sensing for efficient robot behaviour. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI'11)*, 2442–2449.

Hawes, N.; Hanheide, M.; Hargreaves, J.; Page, B.; Zender, H.; and Jensfelt, P. 2011. Home Alone : Autonomous Extension and Correction of Spatial Representations. In *Proc. Int. Conf. on Robotics and Automation*.

Kelleher, J. D., and Costello, F. 2009. Applying computational models of spatial prepositions to visually situated dialog. *Computational Linguistics* 35(2):271–306.

Kelleher, J., and van Genabith, J. 2006. A computational model of the referential semantics of projective prepositions. In Saint-Dizier, P., ed., *Syntax and Semantics of Prepositions*, Speech and Language Processing. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Klenk, M.; Forbus, K.; Tomai, E.; Kim, H.; and Kyckelhahn, B. 2005. Solving everyday physical reasoning problems by analogy using sketches. In *Proceedings of the 20th national conference on Artificial intelligence (AAAI'05)*.

Kuipers, B. 2000. The spatial semantic hierarchy. *Artificial Intelligence* 119:191–233.

Langley, P. in press. Advances in cognitive systems. *AI Magazine*.

Lockwood, K.; Lovett, A.; and Forbus, K. 2008. Automatic classification of containment and support spatial relations in english and dutch. In *Proceedings of the international conference on Spatial Cognition VI: Learning, Reasoning, and Talking about Space*, 283–294. Springer-Verlag.

Regier, T., and Carlson, L. 2001. Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General* 130(2):273–298.

Tellex, S.; Kollar, T.; Dickerson, S.; Walter, M.; Banerjee, A.; Teller, S.; and Roy, N. 2011. Understanding natural language commands for robotic navigation and mobile manipulation. In *Proceedings of the Twenty-Fifth Conference on Artificial Intelligence (AAAI'11)*.

Thrun, S. 2003. Robotic mapping: A survey. In Lakemeyer, G., and Nebel, B., eds., *Exploring Artificial Intelligence in the New Millennium*. Morgan Kaufmann Publishers Inc. 1–35.