The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Mechanical support as a spatial abstraction for mobile robots

Kristoffer Sjöö, Alper Aydemir, Thomas Mörwald, Kai Zhou and Patric Jensfelt

Abstract— **Motivated by functional interpretations of spatial language terms, and the need for cognitively plausible and practical abstractions for mobile service robots, we present a spatial representation based on the physical support of one object by another, corresponding to the preposition "on". A perceptual model for evaluating this relation is suggested, and experiments – simulated as well as using a real robot – are presented. We indicate how this model can be used for important tasks such as communication of spatial knowledge, abstract reasoning and learning, taking as an example direct and indirect visual search. We also demonstrate the model experimentally and show that it produces intuitively feasible results from visual scene analysis as well as synthetic distributions that can be put to a number of uses.**

I. Introduction

The field of service robotics is, at its core, directed toward the creation of systems that are as versatile, adaptive and powerful in everyday environments as human beings are. Only when this becomes true will we be able to depend on robots in the same way as on people around us.

The human machine is superbly adapted to this kind of environment; not just physically (such as having legs for negotiating stairs and thresholds, and arms for opening doors and using appliances), but mentally as well. Human cognition, language, and civilisation have all evolved, and are evolving, in inextricable conjunction with each other. Any cultural or linguistic concept, whether it is the function of a piece of furniture or the meaning of a word, needs the support of cognitive mechanisms; individuals are driven to acquire such mechanisms by reinforcement pressures from their surroundings [8] – while at the same time, the minds of individuals, embodied in the real world, shape and bring forth that same cultural or linguistic concept in turn.

This all suggests the following:

1) Adopting human-like cognitive patterns will help robots approach human-like performance in the context of homes, offices or other environments that are the products of human inclinations, activities and thought.
2) Linguistic concepts can provide insights into cognition that can help understand the nature of those cognitive patterns.

These are the principles on which this work is based. Our research addresses *spatial* concepts specifically. Spatial concepts are of great importance to robotic agents, especially mobile ones:

- They are a necessary part of linguistic interaction with human beings, both when interpreting utterances with a spatial content and when formulating such utterances.
- They allow knowledge transfer between systems, whether different robots, or databases such as the Open Mind Indoor Common Sense database (OMICS) [2] (which contains "commonsense" information about indoor environments provided by humans, such as where objects may be found), as long as those concepts are shared.
- They provide qualitative abstractions that facilitate learning and reasoning.
- They can be used to guide top-down processes such as e.g. visual object search.

Drawing inspiration from results in psycholinguistics, in this paper we examine the functional spatial relation of mechanical support, which in English corresponds to the preposition "on". We contribute a novel and general perceptual measure that allows a robot to analyze a scene in terms of this relation in practice. We implement this perceptual model showing it to produce results in accord with human intuitions of "on"; we also perform simulated sampling experiments to show how it can be used in a top-down fashion to generate a conditional probability distribution over object poses given that the relation is known or assumed to hold.

Other work has examined ways to quantify spatial relations. Inspired by findings on spatial information encoded in the hippocampus, [11] suggests a number of geometrical factors, e.g. coordinate inequalities, that play a part in defining relations such as "below", "near" or "behind", but does not attempt to provide exact formulas.

In [13] the *Attention Vector Sum* is proposed as a practical numerical measure of how acceptable a particular spatial relation is for describing a scene, and this model is compared to actual human responses. The scenes used in this work are 2-dimensional and the trajector (mobile object) is treated as a single point.

[10] presents a system where a user can sketch images of basic figures, and which learns to distinguish between examples of "in", "on", "above", "below" and "left". However, the domain used in the work is strictly 2-dimensional.

Topological relations specifically are surveyed in [3]. *Region connection calculus* and its variants provide a language for expressing qualitative relationships between regions, such as containment, tangential contact etc. Relations are of an all-or nothing nature; and they represent objective, geometrical as opposed to perceptual or functional attributes.

The aforementioned work, because of its emphasis on pure geometry – typically in 2 dimensions – is not directly

K. Sjöö, A. Aydemir and P. Jensfelt are with the Centre for Autonomous Systems at the Royal Institute of Technology (KTH), Stockholm, Sweden. T. Mörwald and K. Zhou are with the Automation and Control Institute, Vienna University. This work was supported by the SSF through its Centre for Autonomous Systems (CAS), and by the EU FP7 project CogX and the Swedish Research Council, contract 621-2006-4520 (K. Sjöö)

suited for applications in a practical mobile robotic scenario. This paper, in contrast, takes a novel, functional approach by basing a relation on a single fundamental, objective mechanical property. Another contribution lies in treating all the objects as entire bodies rather than simplifying them into points, a simplification which ignores the importance of physical contact in the "on" relation. We also show how the method can be used to generate probability distributions, such as might be used for visual search.

This paper is organized in the following way: Section II introduces the spatial relation we are examining and our suggested perceptual model for it; Section III presents the implementation of the model that we have carried out and the experiments performed – on real image data as well as simulated. Section IV discusses the results and directions for future research, followed by conclusions in Section V.

II. The On relation

Spatial predicates in language come in different categories. *Projective* spatial relations constrain the trajector's[1] location within an essentially *directed* region relative to the landmark. Examples in English include "to the left of", "behind" and "past". *Topological* relations, in contrast, locate the trajector in some manner that is independent of direction. Typical examples are "on", "at" and "inside". Topological relations seem to be among the first to be learned in humans [12]. In this work, we are concerned with "on", an important English word implying an equally important underlying spatial concept.

Research suggests that verbal descriptions of space do not, in general, correspond one-to-one to cognitive representations [9]. Instead, it seems conceptualization forms around kernels of *functional* criteria, such as "physical attachment", "superposition" (an object being located in the space vertically above another) or "containment" (an object being enclosed by another). As has been noted by e.g. Talmy [15] and Herskovits [6], English' "on" carries a central meaning also represented in many other languages: that of *support* against gravity; i.e., a trajector is "on" a landmark if it would, were the landmark to be removed, begin to fall or move under the influence of gravity. This sense of "on" is an *idealized cognitive model* or ICM [7], around which other, less central and more idiomatic senses of "on" form in a way specific to each language.

A. The importance of support in robotics

We observe that the notion of support is highly related to the functional aspects of space as designed, constructed and lived in by human beings. Such space is full of entities specifically made to support others, both statically – such as tables, shelves, counters, chairs, hooks and desks – and dynamically – such as trays, trolleys, and dishes. This functional aspect is emphasized by Coventry and Garrod [4]:

> Describing where an object is located goes beyond the description of a geometric position of objects as a snapshot in time. Understanding spatial language is also about the *purpose* that location serves for the users of that language.

As for "on", it is the 14th most common English word [1] which indicates the importance that humans attach to support in representing the spatial location of an object[2].

Apart from the evidence given by its prominent role in language (and thus in the minds of people), support is an intuitively useful abstraction in the following way: If a support is moved, then supported objects will tend to move with it, maintaining the relation (Coventry and Garrod refer to this as "Location control" [4]), and it makes the relation inherently hierarchical, which is a useful property in spatial organization.

Also, the fact that artifacts in the environment are explicitly designed to provide support surfaces for objects means that often, when an object is "on" another, it *belongs* there functionally to some degree and is thus likely to be replaced on the same surface even after a human picks up, manipulates, or moves it – notwithstanding that the exact position may have changed. For example, a desk may be shifted or moved, or worked at by its owner, and its set of supported objects yet be unchanged.

It thus is of interest to robotics to use a spatial representation that encodes this functional relationship between objects. Although this work is inspired by linguistic clues, giving a robot additional linguistic capabilities is only an incidental outcome. It is also necessary to point out that the word "on" spans far more meanings than the core physical support relation: it may entail indirect rather than direct support, adhesion or suspension, as well as metaphorical meanings. Here, we are not attempting to cover that complexity.

B. A perceptual model

The "support" relation proposed above constitutes an idealized model, but is as such not possible to evaluate directly from perceptual data. Neither robots nor humans can ascertain degree of mechanical support merely by visually regarding a scene, and so it becomes necessary to introduce a perceptual model to estimate the ideal relation.

Humans use context, experience with specific objects and generalizations, as well as schemata, to decide whether an object is "on" another. For robots, we model this with a simplified 3-dimensional geometric predicate, termed On, such that On(A, B) corresponds to "A is supported by B". The relation is graded and can attain values in the range $[0, 1]$.

The following are our criteria and their justification. O denotes the trajector object, and S the support object or landmark. The criteria are illustrated in Figure 1.

[1]The trajector is the entity whose location (and/or motion) is being denoted explicitly, in relation to the landmark. Thus, in the sentence "A is above B", A is the trajector and B the landmark.

[2]Though many usages of "on" in English are not about support directly, or even about physical space, the fact that "on" is the word used still underscores the cognitive centrality of its core meaning.

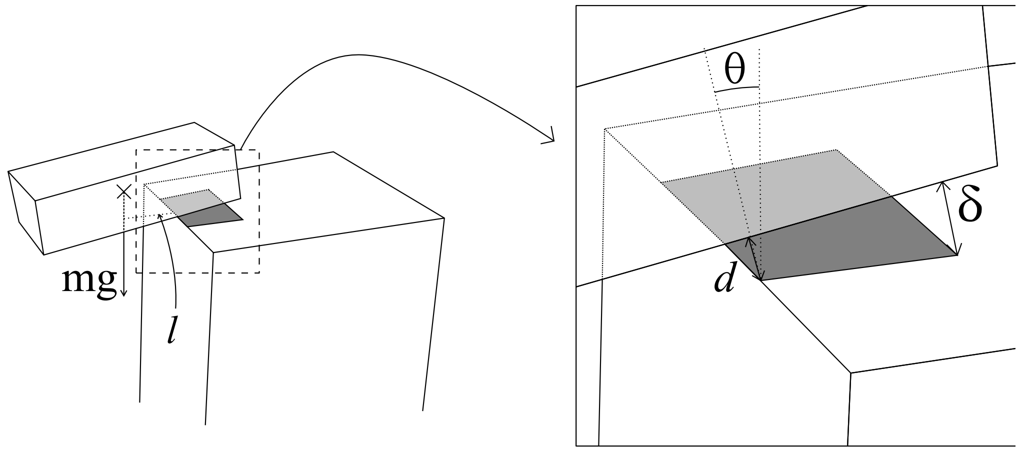

Fig. 1. Key features used in computation of ON: Separation d, COM offset l, contact angle θ and contact threshold δ. The gray area represents the contact.

1) *Separation between objects*, d. d can be positive or negative, negative values meaning that objects seem to be interpenetrating.
   In order for an object to mechanically support another, they must be in contact. Due to imperfect visual input and other errors, however, contact may be difficult to ascertain precisely. Hence, the apparent separation is used as a penalty.
2) *Horizontal distance between COM and contact*, l. It is well known that a body O is statically stable if its center of mass (COM) is above its area of contact with another object S; the latter object can then take up the full weight of the former. Conversely, the greater the horizontal distance between the COM and the contact, the less of the weight S can account for, as the torque gravity imposes on O increases, and this torque must be countered by contact with some other object.
   Thus we impose a penalty on $\textsc{On}(O, S)$ that increases with the horizontal distance from the contact to the COM of O. The contact is taken to be that portion of S's surface that is within a threshold, δ, of O, in order to deal with the uncertainties described above. If $d > \delta$, the point on S closest to O is used instead. l is the positive distance to the outer edge of the contact area if outside it, and the negative distance if inside.
3) *Inclination of normal force*, θ – the angle between the normal of the contact between O and S on the one hand, and the vertical on the other. The reason for including this is that *mutatis mutandis*, the normal force decreases as the cosine of θ, meaning the weight of O must be either supported by another object or by friction (or adhesion).

All these values can be computed from visual perception in principle. Unless otherwise known in advance, the position of the COM is taken as the geometrical centroid of the object (since density cannot be determined by vision).

In order to allow a measurable value to be computed, the agreement with each of the three above criteria is represented as a continuous function, with a maximum at the point of best agreement with the criterion. This provides robustness against error. Criterion 1 is represented by an exponential *distance factor*:

$$\textsc{On}_{distance}(O, S) \triangleq \exp\left(-\frac{d}{d_0(d)} \ln 2\right) \tag{1}$$

where d_0 is the falloff distance at which ON drops by half.

$$d_0 = \begin{cases} -d_0^-, & d < 0 \\ d_0^+, & d >= 0 \end{cases}$$

The constants d_0^- and d_0^+ are both greater than 0 and can have different values (representing the penetrating and nonpenetrating cases, respectively).

Criteria 2 and 3 make up the sigmoid-shaped *contact factor*:

$$\textsc{On}_{contact}(O, S) \triangleq \cos\theta \cdot \frac{1 + \exp(-(1 - b))}{1 + \exp\left(-\left(\frac{-l}{l_{\max}} - b\right)\right)} \tag{2}$$

Here, $l_{\max}$ is the maximum possible distance an internal point can have within the contact area, and b is an offset parameter.

The values are combined by choosing whichever factor is smaller, indicating the greater violation of the conditions for support:

$$\textsc{On}(O, S) \triangleq \min(\textsc{On}_{contact}, \textsc{On}_{distance}) \tag{3}$$

Note that the resultant value of ON, although in the range $[0, 1]$, is not a probability. Rather, it represents the degree of resemblance of the visual scene to the prototypical ON case. It can be thresholded to produce a true/false judgement, which may in turn be utilised in a qualitative reasoning framework, or for learning – such as learning relationships between object types in an environment. Alternatively, the ON measure could be compared with similar measures for other relations or other objects, to determine which linguistic description of the scene is the most apt. It can also be used to weight samples to produce a distribution over poses of O, as discussed below.

C. Probability modelling

The conceptualization above does not explicitly make use of any probabilities. However, it is obvious that the fact of an object being ON another is not sufficient to recover the exact pose of the trajector. A probability distribution over poses can be produced in the following way:

Given the pose and geometry of the landmark S, and the geometry (but not the pose) of the trajector O, each possible pose π for the trajector yields a value of $\text{ON}(O_\pi, S)$ for that pose.

It is now possible to introduce probabilities in the following way. Introduce a true/false event $On(O, S)$ signifying that $\text{ON}(O, S) > t$ where t is a threshold. Then,

$$p(\pi|On(O_\pi, S)) = \frac{p(On(O_\pi,S)|\pi)p(\pi)}{p(On(O_\pi,S))} = \\ = \frac{[\text{ON}(O_\pi,S)>t]p(\pi)}{p(On(O_\pi,S))} \tag{4}$$

Here [] denotes the Iverson bracket:

$$[X] = \begin{cases} 1, & \text{if } X \text{ is TRUE} \\ 0, & \text{otherwise} \end{cases}$$

In other words, the probability is simply proportional to the prior for the pose π whenever $\text{ON}(O_\pi, S) > t$, and 0 elsewhere. Though it may be hard to express this distribution analytically, by drawing samples randomly from $p(\pi)$, discarding those failing to reach the threshold, and normalising over the remainder, an arbitrarily good approximation can be found.

D. Example: Visual object search

One use for the above probabilistic formulation is the task of locating an object by searching for it visually [16], [18], [19]. Visual object search is typically posed as the problem of selecting a series of views $\{V_i\}$, such that the cost of acquiring and processing those views is minimized while detecting the sought object at some set probability.

Assume that some algorithm exists that produces a sequence of views, given a probability distribution for the sought object $p(\pi_O = x) = f_O(x)$, the views incurring the total cost $C_O\{f_O\}$. The cost may depend on the actual object, due to size, saliency et cetera.

In this context, the ON relation can be highly useful. In many scenarios, the exact position of an object O may be uncertain or unknown, even while it is known or presumable that it is ON some other object S. This information can have several sources: O may have been seen ON S at an earlier time, and location control implies the relation will still hold even if S has moved. The connection may also be statistical in nature, learned through experience from many analysed scenes ("this type of object is usually located ON that type") or from a commonsense knowledge database. The information may also come from symbolic reasoning or linguistical utterances.

Using an object's location to help search for another is known as *indirect search*. Indirect search was first investigated in 1976 by Garvey [5]; there, a system looking for a phone in a room is first tasked with finding the table that the phone is resting on. Wixson [17] re-visited the idea of indirect search in the context of mobile robotics; however, previous work on exploiting spatial relations to guide the visual search process on mobile robots is non-existant.

If it is known a-priori that $On(O, S)$, and the location of S is known, then the above distribution may be used as a prior probability input to a view-selection algorithm, at cost $C_O\{f_{O|S}\}$.

If $On(O, S)$ is known to hold but S' location is not known, there are two choices: Indirect search can be used, i.e. locating S first and then locating O given the position of S. The cost of this will be[3]:

$$C_S\{f_S\} + C_O\{f_{O|S}\}$$

Alternatively, one may use a distribution over O's location obtained through chain inference:

$$C_O\{f_O\} = C_O\left\{\int_S f_{O|S} f_S\right\}$$

Either approach can be evaluated using the sampling method suggested above. By comparing the costs, the most beneficial option can be selected depending on the situation.

III. EXPERIMENTS

To test the feasibility of the concepts described in the preceding section, we have implemented them on a robotic system and tested it in a real-world setting.

We also present a series of simulations that illustrate the potential of the approach using random sampling to synthesize a distribution over positions in space.

A. Experimental setup

The robot used in our experiments is a Pioneer III wheeled robot, equipped with a stereo camera mounted on a pan-tilt unit at 1.4 m above the ground.

Three different box-shaped objects were used for the tests: A, B and C, as seen in Figs. 2–5. Objects were detected and an initial pose estimated using SIFT features, and the pose refined and tracked using particle filtering based on edge information acquired from the known geometric model of each object [14]. Furthermore, horizontal plane patches were extracted from stereo depth information and assembled into planar objects (table surfaces).

The resulting object poses, along with their known geometries, were then processed by the ON computation described in Section II, using the parameter settings δ =3 cm, d_0^+ =2 cm, d_0^- =1.4 cm and $b = 0.5$. The center-of-mass of each object was taken to be its geometrical center.

B. Results

Figure 2 shows a simple case (The wireframe contours show the estimated object poses output by the tracking algorithm). The values for the support function in this scene are:

[3] Although the second term cannot be known exactly without knowing S' orientation, one can compute an average over orientations or use a typical orientation; either way, the cost will not vary much for most objects.

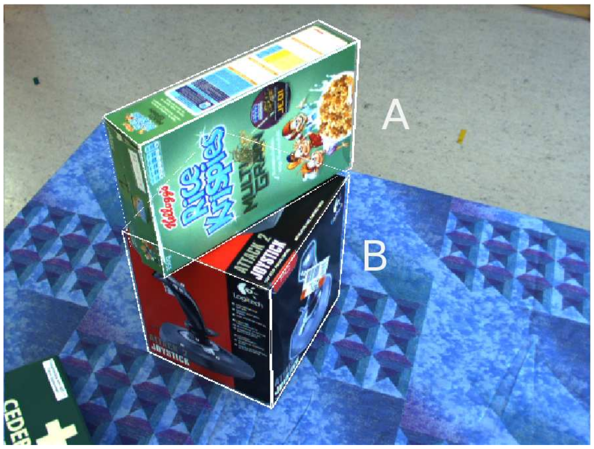

Fig. 2. Typical case: A ON B, B ON table

Fig. 3. Ambiguous case: B partly ON A, B partly ON table

	$\text{ON}(A, x)$	$\text{ON}(B, x)$
$x = A$	—	0%
$x = B$	93%	—
$x =$Table	2%	92%

The support relation is unambiguous in this case: A is supported by B, and B by the table. In Figure 3, the situation is more ambiguous, with B resting partly on the table, and also leaning on A. The ambiguity is reflected in the ON measures:

	$\text{ON}(A, x)$	$\text{ON}(B, x)$
$x = A$	—	25%
$x = B$	0%	—
$x =$Table	74%	47%

Figure 4 shows another double support example; the object is held up approximately equally by the two objects, which is reflected in the computed function:

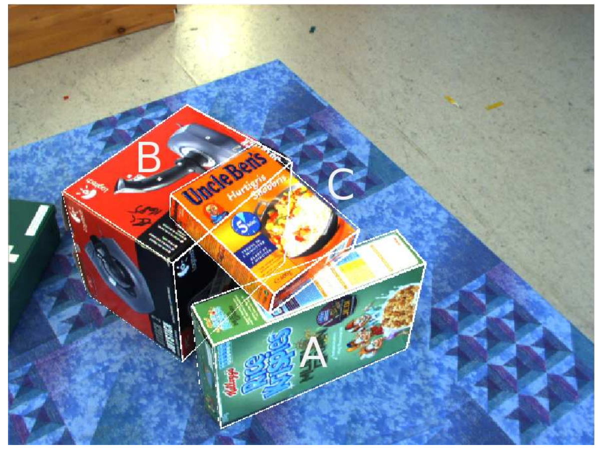

Fig. 4. Ambiguous case: C partly ON each of A and B

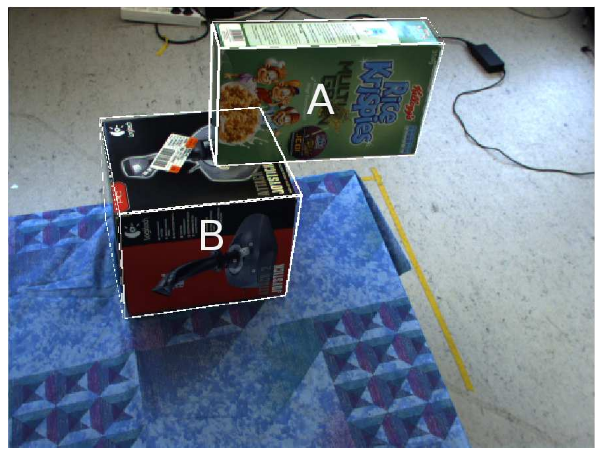

Fig. 5. An anomalous case

	$\text{ON}(A, x)$	$\text{ON}(B, x)$	$\text{ON}(C, x)$
$x = A$	—	1%	28%
$x = B$	0%	—	30%
$x = C$	0%	0%	—
$x =$Table	91%	93%	3%

Finally, Figure 5 depicts a situation that is seemingly physically implausible.

	$\text{ON}(A, x)$	$\text{ON}(B, x)$
$x = A$	—	0%
$x = B$	22%	—
$x =$Table	4%	84%

The ON measure here is low, and even though there is no other object with which to compare it, the low value means the configuration is far from prototypical and not one that would be expected by the robot, given only the information that "A is on B". The problem here is that the COM has been modified with an extra weight so as not to be at the geometrical center of A, but the robot doesn't know this, and as stated earlier it cannot be gleaned from vision alone.

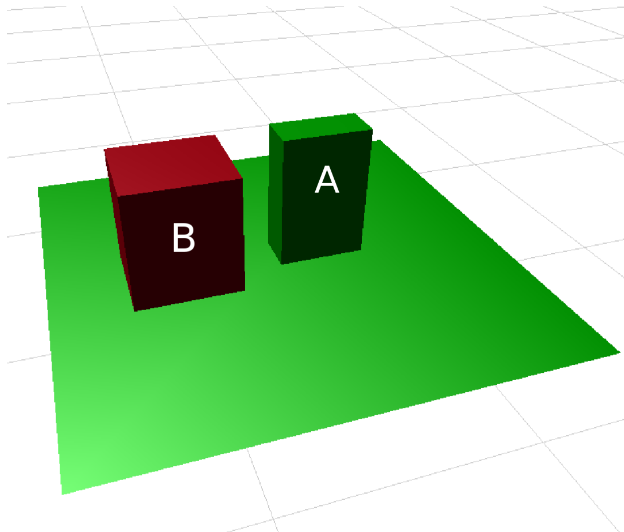

Fig. 6. Objects used in simulation experiments

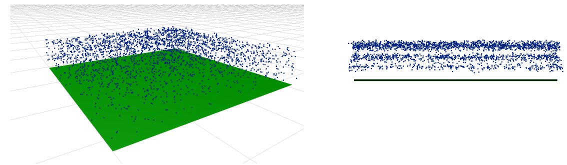

Fig. 7. Position of A, given "A On Table"

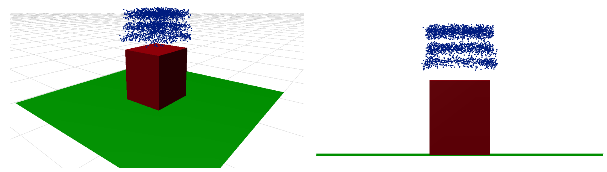

Fig. 8. Position of A, given "A On B"

In summary, we have verified that our approach works in an implemented real-world system, all the way from sensors to spatial abstraction, producing outputs that are intuitively reasonable.

C. Simulation

For the simulated experiments, we used the same object geometry models as in the real-life experiments, as shown in Figure 6. One or more objects were fixed to one position in space (considered "known"), and one or more objects were assigned variable poses (considered "unknown"). Because of the lack of noise, we were able to use stricter parameter settings: δ =2 cm, d_0^+ =0.7 cm, d_0^- =0.4 cm and $b = 0.5$.

In accordance with the principles put forth in Section II-C, we then sampled the distribution of the On function by randomly selecting poses for the variable objects and evaluating the On function for each. The figures in this section each show 2500 samples that evaluated to On > 0.5. Note that the full 6 DOF pose was variable, although the figures only show the position of the COM.

Two basic cases are shown in Figures 7 and 8. The former shows samples of A's position, given that it is On the table; the latter, given that it is On B. The stratification that can be observed corresponds to A standing up, and lying on its side or back, respectively. This arises directly from the On function and shows how On can encapsulate complex modes of configurations implicitly. Automatic clustering would allow for the extraction of these modes, which might then be used in high-level qualitative reasoning.

Two other configurations of the object B are shown in Figure 9. These illustrate how the inclination of the support object is taken into account in the On computation. Not all points "above" B are valued equally, as might be the case in a purely geometrical approach, but rather points corresponding to a largely vertical contact normal are considered more feasible. In the second image, the distribution is concentrated to a narrow region corresponding to A balancing on the topmost edge of B (which translates to a low On value in absolute terms, despite being the global maximum).

The potential uses of these distributions are many. As explained in Section II-C, they can be translated into probability distributions. In a search scenario, where it is known that A On B, and B's pose is known, the distribution may be used to direct the search for A. If the pose of B is not known, the distribution (as computed by assuming B *were* known) can be compared to an uninformed prior on A's location, allowing the robot to decide whether it is worth it to search for B first, or if it is better to look for A directly.

In that vein, Figure 10 contains the result of a *chained* sampling, where both objects A and B were allowed to vary randomly. Only when both B On Table and A On B were greater than 0.5 was the position of A plotted. In other words, what is represented is the distribution over A's position, given that A On B and B On Table, but with B's exact pose unknown.

This type of chained inference allows for e.g. searching for A without first locating B, while still utilizing the knowledge that A On B. As stated above, the distribution can be compared to the prior of A, and A given A On B, to determine whether it is more beneficial to locate B first or not.

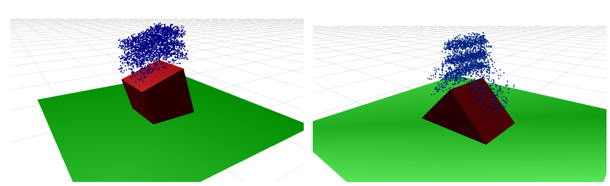

Fig. 9. Position of A, given "A On B"

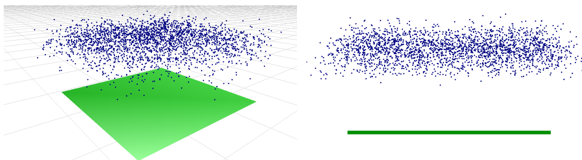

Fig. 10. Position of A, given "A On B, B On Table"

IV. Discussion

This work has only begun to explore the possible uses to a mobile robot of the conceptualization proposed. We would like to extend the work in several ways. First, evaluating the efficiency of object search utilising the results of this paper, as well as exploring how the principle of using functional criteria can be generalized through a similar treatment of other important spatial relations, especially topological ones such as "in" or "at". In addition, we hope to integrate this conceptualization of "on" with information from language and the OMICS database, and conversely to use it to generate spatial utterances and generalizations about typical relations between objects. Furthermore our implementation has been limited to box and plane shapes; it will readily extend to any convex 3D shapes, but non-convex shapes require that further assumptions be made.

The perceptual model described in Section II-B assumes knowledge of the involved objects' geometry, poses and centers of mass. Whereas a human is able to estimate these quantites, even for novel objects, and/or extrapolate them based on experience, a robot may not always have access to good such estimates from its visual system. Vision is not the focus of this work, however, and the soft nature of the functions gives some robustness to poor visual information; moreover, and more importantly: as shown in Section II-D, when the relation information is used in the "synthesis" direction, such as in search, poses do not need to be provided.

The descriptors used in this work were selected in an *a priori* fashion, and the relevant weights were adjusted manually. In the future, we hope to achieve a more objective correspondence between the model and the idealized conceptualization of support – and other such idealized conceptualizations – through learning, either based on studies of human classification or on mechanical simulation. The choice of which features to use in the first place is a still more challenging learning goal, which must nevertheless be tackled in order to allow the approach to be applied to a far wider set of conceptualizations. A concept such as mechanical support cannot be acquired from scratch without learning from experience with manipulating physical bodies, connecting physical forces that are felt to visual properties of objects and the effects of actions. More work needs to be done on using such feedback to build functional spatial concepts, work that cannot be separated from the larger scope of imbuing robots with greater intelligence.

V. Conclusions

We proposed an idealized cognitive model for the core concept underlying English "on", *viz.* mechanical support, in order to give us a functionally grounded abstraction primitive to use with qualitative reasoning and learning, top-down processes such as visual search, and linguistic interaction. A novel perceptual model was designed and implemented to approximately analyze real-world scenes in terms of this model, and results of experiments with real-world data were presented. Finally, we contributed a method to synthesize expectations about the metric location of an object to aid in e.g. efficient search.

References

[1] Askoxford: Language facts. http://www.askoxford.com/oec/mainpage/oec02/, 2010.

[2] Openmind indoor commonsense. http://openmind.hri-us.com/, 2010.

[3] A.G. Cohn and S.M. Hazarika. Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 2001.

[4] Kenny Coventry and Simon Garrod. *Saying, seeing and acting: the psychological semantics of spatial prepositions*. Hove, 2003.

[5] Thomas David Garvey. *Perceptual strategies for purposive vision.* PhD thesis, Stanford, CA, USA, 1976.

[6] A. Herskovits. *Language and Spatial Cognition*. Cambridge University Press, 1986.

[7] G. Lakoff. *Women, fire and dangerous things: what categories reveal about the mind.* University of Chicago Press, 1987.

[8] S. Levinson. *Language and Space*, chapter Frames of Reference and Molyneux's question: cross-linguistic evidence. MIT Press, 1996.

[9] S. Levinson and S. Meira. 'natural concepts' in the spatial topological domain – adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*, 79(3), 2003.

[10] K. Lockwood, K. Forbus, D.T. Halstead, and J. Usher. Automatic categorization of spatial prepositions. In *Proceedings of the 28 th Annual Conference of the Cognitive Science Society.*, 2006.

[11] J. O'Keefe. *The Spatial Prepositions*, chapter 7. The MIT Press, 1999.

[12] J. Piaget and B. Inhelder. *The Child's Conception of Space*. Routledge & Keagan Paul Ltd., 1956.

[13] T. Regier and L. A. Carlson. Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology*, 130(2):273–2098, 2001.

[14] A. Richtsfeld, T. Mörwald, M. Zillich, and M. Vincze. Taking in shape: Detection and tracking of basic 3d shapes in a robotics context. In *Computer Vision Winter Workshop*, pages 91–98, 2010.

[15] L. Talmy. Force dynamics in language and cognition. *Cognitive Science*, 1988.

[16] John K. Tsotsos and Ksenia Shubina. Attention and visual search : Active robotic vision systems that search. In *International Conference on Computer Vision Systems ICVS'07*, page 539, Washington, DC, USA, 2007. IEEE Computer Society.

[17] Lambert E. Wixson and Dana H. Ballard. Using intermediate objects to improve the efficiency of visual search. *Int. J. Comput. Vision*, 12(2-3):209–230, 1994.

[18] Yiming Ye and John K. Tsotsos. Where to look next in 3d object search. In *ISCV '95: Proceedings of the International Symposium on Computer Vision*, page 539, Washington, DC, USA, 1995. IEEE Computer Society.

[19] Yiming Ye and John K. Tsotsos. Sensor planning for 3d object search. *Comput. Vis. Image Underst.*, 73(2):145–168, 1999.