I. INTRODUCTION

Ongoing cognitive development during the first years of human life may be the result of a set of developmental mechanisms which are in continuous operation [1]. One such mechanism identified is the ability of the developing child to learn effective preconditions for their behaviours. It has been suggested [2] that through the application of behaviours involving more than one object, infants begin to learn about the relations between objects.

We consider a precondition to be a learnt decision rule by which some features of the environment are used to predict the successful outcome of a behaviour. This can be used as a planning operator to allow a robot to sequence learnt actions to achieve a goal. The limited scope of this definition allows us to approach the problem computationally. This concept of a precondition is loosely related to the notion of an affordance [3] used as a planning operator, which has been well studied within the field of developmental robotics (see e.g. [4], [5]).

Learning a precondition for a motor action from raw sensor data is challenging as it may take many thousands of examples to learn an effective rule. For this reason we first perform an abstraction to convert data into a form which simplifies the learning procedure.

In this work, we learn a limited number of abstractions which can then be used to form preconditions for motor actions. These abstractions take the form of spatial relations amongst objects. We consider three “classes” of spatial relation: The objects either are separated from, on-top of, or inside each other. We have tackled this same problem in previous work [6]. Here we report on recent improved results using a novel application of histograms to visually recognise a spatial relation between objects in the environment. Using this histogram based approach we are able to report a very high rate of success when the system is asked to recognise a spatial relation.

II. LEARNING SPATIAL RELATIONS

For learning spatial relations between objects we used data from a sophisticated 3D vision system inside the physically realistic simulator RobWorkSim [7]. In our experiments we use 4 different household objects (see Figure 1a). Using these objects we are able to design combinations of object pairs accounting for each of the three classes of spatial relation.
We create 2D histograms which store relevant information about the spatial relations between objects. To extract this information from the sets of texlets, we calculate two distance measures between all texlets of one object to all texlets of another object: The absolute distance of a pair of texlets along the $xy$ plane and the relative distance of the first object’s texlets to the second object’s on the $z$-axis, such that if the first texlet is above the second texlet, the distance is positive, otherwise negative. Since we always consider pairs of objects, if the first object has $n$ texlets and the second has $m$ texlets, this results in a $nm$ vector for each measure considered. These distances are used to fill the histograms, such that the the $x$-axis is the absolute $xy$ distance and the $y$-axis of the histogram is the relative distance between texlets along the $z$-axis. The histograms have 50 bins per axis, experientially this value produced the best results. The $x$-axis runs from 0mm to 2000mm, and the $y$-axis from $\pm 150$mm. Example histograms are given in Figure 2.

For learning to differentiate the spatial relations between objects, we used a Random Forest (RF) classifier [10]. The RF had 100 trees, an overlap of 0.9 and 400 inputs per tree. For each class we used 5400 samples to build a training set. The training set contained samples from all configurations of each class in equal numbers (e.g. from on-top there were 2700 sample from coffee cup and tray and 2700 from the turtle and tray.) Similarly, in the validation set each class was represented equally with 1404 samples per class$^1$. After training, the system classified 99.95% of the 4212 validation samples correctly. Although our result is not directly comparable, we achieved a higher success rate than [11] with similar data.

We tested the system on its ability to generalise to novel objects: We introduced a 6-sided die either next to or on-top of wedge. The system performed well, classifying 95% of the samples correctly. Results are shown in Table II.

III. DISCUSSION AND RELATED WORK

The most closely related work on learning spatial relations between objects in a 3D space is [11] who use a support vector machine based approach. In this approach the support vectors are picked from for their ability to differentiate the point cloud into two objects. This has the effect that the subset of points considered by the classifier are on the edges of the object. Relations are then learnt based upon the relative positions of clusters of the support vectors. For any classification based approach to be successful, it requires that similar relations have a similar representation; at the level of point clouds/texlets the representation of a relation can be very different. In the case of [11], the scene is reduced to clusters with $xyz$ coordinates. We feel that our histogram based approach allows for a more generic representation of the scene — we maintain a higher proportion of the important information about the relations between objects.

ACKNOWLEDGMENT

This work was supported by the EU Cognitive Systems project XPERIENCE (FP7-ICT-270273) and Leverhulme Grant F/00 152/AL.

REFERENCES


---

$^1$For some classes, the object pairings were not equally represented.

---

**TABLE I: Known objects**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Separated</td>
<td>1404</td>
<td>1</td>
<td>2807</td>
<td>0</td>
</tr>
<tr>
<td>On-top</td>
<td>1404</td>
<td>0</td>
<td>2808</td>
<td>0</td>
</tr>
<tr>
<td>Inside</td>
<td>1403</td>
<td>0</td>
<td>2808</td>
<td>1</td>
</tr>
</tbody>
</table>

**TABLE II: Novel objects**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Separated</td>
<td>1421</td>
<td>105</td>
<td>1347</td>
<td>31</td>
</tr>
<tr>
<td>On-top</td>
<td>1262</td>
<td>30</td>
<td>1422</td>
<td>190</td>
</tr>
<tr>
<td>Inside</td>
<td>0</td>
<td>86</td>
<td>2807</td>
<td>0</td>
</tr>
</tbody>
</table>