Probabilistic Representation of Objects and their Support Relations

Fabian Paus and Tamim Asfour

Institute for Anthropomatics and Robotics,
Karlsruhe Institute of Technology (KIT),
Adenauerring 2, 76131 Karlsruhe, Germany
{paus,asfour}@kit.edu

Abstract. Understanding uncertainty about objects and their relations in a scene is essential for action selection in robotics. We propose a novel approach for a probabilistic representation of objects and their support relations taking into account pose and shape uncertainty. Starting with a segmented point cloud a probability distribution over the object geometry is estimated, from which samples are drawn to calculate a probability distribution over support relations. To evaluate the approach, we created a new RGB-D dataset, the KIT Support Relation dataset (KIT-SR), consisting of 60 scenes annotated with pixel-wise object labels and ground-truth support relations. Furthermore, we augmented the Object Segmentation Database (OSD) with support relation annotations. We evaluated our proposed probabilistic approach against two state-of-the-art deterministic approaches and show significantly improved precision, recall, F1, and Brier scores.

Keywords: Uncertainty Representation · Object Relations · RGB-D

1 Introduction

In cluttered scenes, humans are able to intuitively understand which objects are supported by each other and which objects can be safely removed without causing the scene to collapse [2]. To equip a robot with such an intuitive physics understanding, support relations between objects need to be extracted from its sensors. Pure computer vision approaches have been developed to extract support relations between segmented image parts using MAP (Maximum a posteriori) inference [8], energy function minimization [3,10], or stability reasoning [13]. The goal of these approaches is to exploit support relations to improve image segmentation. In robotics, support relations are required to determine manipulation order in clutter [6]. Furthermore, the extracted relations should be physically plausible, enabling the robot to select actions [5] and predict action outcomes [4]. These approaches rely on the robot's knowledge about the camera pose, and therefore the gravity vector. By determining which objects exerts a force on another, support relations are derived. Recently, different learning algorithms using CRFs [11] and CNNs [12] have been proposed to extract manipulation order directly from image data.

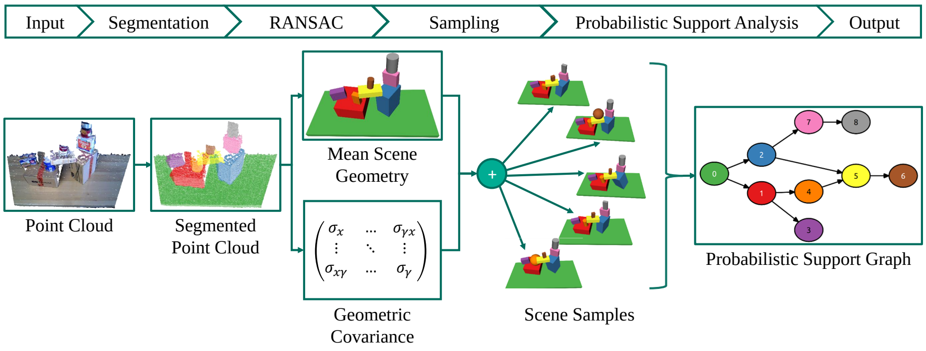

Fig. 1. Overview of the probabilistic analysis of support relations based on point clouds.

While the approaches from the computer vision community have leveraged probabilistic representations for support relations (e. g. [8,12]), the approaches from the robotics community are still limited to deterministic representations (e. g. [5,4]). In this work, we address this research gap by investigating the benefits of a probabilistic representation for support relations to determine manipulation order in cluttered scenes.

In our previous work, we developed a method for extracting physically plausible support relations between unknown objects based on RGB-D images [4]. We used geometric primitive fitting based on RANSAC (Random Sample Consensus) to estimate the pose and shape of unknown objects in the scene. However, this approach did not consider the uncertainty associated with the pose and geometry estimation. Therefore, we will present a novel probabilistic scene representation that models both pose and shape uncertainty. By sampling from the resulting geometric scene distribution, we determine a probability distribution over support relations in the scene. We show that this approach improves the detection rate of support relations significantly and enables the humanoid robot ARMAR-6 [1] to manipulate cluttered table-top scenes safely.

2 Technical Approach

First, we present a novel probabilistic scene and support relation representation, which explicitly models uncertainty in object geometry and support relation existence. Then, we show how to determine the parameters of this representation based on a point cloud of the perceived scene. Fig. 1 shows an overview of our approach.

2.1 Probabilistic Scene and Support Relation Representation

We assume that the shape of an object in a scene can be approximated using geometric primitives. In this work, a geometric primitive can be a box, a cylinder, or a sphere, but the set of primitives can be trivially extended. Each primitive $p_i = (t_i, \mathbf{x}_i)$ consists of a primitive type $t_i \in \{\text{Box}, \text{Cylinder}, \text{Sphere}\}$ and a state vector $\mathbf{x}_i \in \mathbb{R}^{N(t_i)}$. The state vector parametrizes the geometry of a primitive, e. g. a sphere is parameterized by four variables: a three-dimensional center point and a radius. The size $N(t_i)$ of the state vector depends on the primitive type:

- $N(\text{Box}) = 10$ (Position, Orientation, Extents)
- $N(\text{Cylinder}) = 8$ (Position, Direction, Radius, Height)
- $N(\text{Sphere}) = 4$ (Position, Radius)

A scene consists of a set of n objects $\mathcal{O} = \{o_1, o_2, \dots, o_n\}$. The geometry of each object o_i is represented as a joint probability distribution $P(t_i, \mathbf{x}_i \mid o_i)$ over the primitive type t_i and state vector $\mathbf{x}_i$.

$$P(t_i, \mathbf{x}_i \mid o_i) = P(t_i \mid o_i) \cdot P(\mathbf{x}_i \mid t_i, o_i)$$

Since the primitive type and state vector are dependent variables, we can represent the joint probability distribution as the product of a discrete distribution over primitive types $P(t_i \mid o_i)$ and a dependent distribution over state vectors $P(\mathbf{x}_i \mid t_i, o_i)$. We choose a multi-variate Gaussian distribution to represent $P(\mathbf{x}_i \mid t_i, o_i)$:

$$\begin{aligned} P(\mathbf{x}_i \mid t_i = \text{Box}, o_i) &\sim \mathcal{N}(\mu_{i,\text{Box}}, \Sigma_{i,\text{Box}}) \\ P(\mathbf{x}_i \mid t_i = \text{Cylinder}, o_i) &\sim \mathcal{N}(\mu_{i,\text{Cylinder}}, \Sigma_{i,\text{Cylinder}}) \\ P(\mathbf{x}_i \mid t_i = \text{Sphere}, o_i) &\sim \mathcal{N}(\mu_{i,\text{Sphere}}, \Sigma_{i,\text{Sphere}}) \end{aligned}$$

Note that the mean $\mu_{i,t_i} \in \mathbb{R}^{N(t_i)}$ and the covariance $\Sigma_{i,t_i} \in \mathbb{R}^{N(t_i) \times N(t_i)}$ have different dimensions corresponding to the primitive type and its state vector dimension.

Drawing a sample from the joint distribution $P(t_i, \mathbf{x}_i \mid o_i)$ produces a concrete geometric primitive $p_i^s = (t_i^s, \mathbf{x}_i^s)$. To implement this sampling, first, a primitive type t_i^s is sampled from the discrete probability distribution $P(t_i \mid o_i)$. Then, a state vector $\mathbf{x}_i^s$ is generated by sampling from the multivariate Gaussian distribution $\mathcal{N}(\mu_{i,t_i^s}, \Sigma_{i,t_i^s})$. Note that the primitive type t_i^s is a concrete value and not a random variable.

Given a set of objects $\mathcal{O}$, we represent the probability distribution over the complete scene geometry as a joint distribution over independent object geometries. Therefore, we can express the geometric scene distribution as the product of distributions for the set of primitives $\mathcal{P} = \{p_1, p_2, \dots, p_n\}$:

$$P(\mathcal{P} \mid \mathcal{O}) = \prod_{i=1}^{n} P(p_i | o_i) = \prod_{i=1}^{n} P(t_i, \mathbf{x}_i | o_i)$$

In this work, we analyze which objects are physically supported by other objects in order to determine safe manipulation sequences. A support relation $\texttt{SUPP} \subseteq \mathcal{O} \times \mathcal{O}$ is a binary relation between objects. We define that a support relation $\texttt{SUPP(A,B)}$[1] between objects $\texttt{A} \in \mathcal{O}$ and $\texttt{B} \in \mathcal{O}$ exists if and only if

[1] We use the notation $\texttt{SUPP(A,B)}$ for $(\texttt{A},\texttt{B}) \in \texttt{SUPP}$.

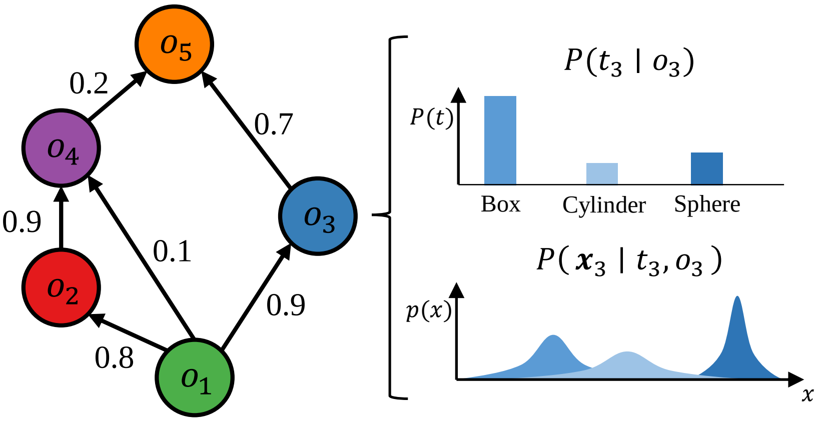

Fig. 2. A probabilistic support graph for the object set $\mathcal{O} = \{o_1, o_2, o_3, o_4, o_5\}$. Each vertex contains the parameters for distributions over primitive type and state vector. Each edge is annotated with the existence probability of a support relation between the corresponding objects.

removing $\mathtt{A}$ causes $\mathtt{B}$ to lose its motionless state [5]. The existence probability of a support relation $\mathtt{SUPP}(\mathtt{A},\mathtt{B})$ depends not only on the objects $\mathtt{A}$ and $\mathtt{B}$, but also on other objects in $\mathcal{O}$:

$$P(\mathtt{SUPP}(\mathtt{A},\mathtt{B}) \mid \mathcal{O}) = P((\mathtt{A},\mathtt{B}) \in \mathtt{SUPP} \mid \mathcal{O}), \quad \mathtt{A},\mathtt{B} \in \mathcal{O}$$

We can now represent a scene and the corresponding support relations as a directed graph $\mathcal{G} = (V, E)$, in which the vertices $V = \mathcal{O}$ represent objects and the edges $E = \mathtt{SUPP}$ represent support relations. By considering the probability distributions over scene geometry and support relations, we can create a probabilistic support graph $\mathcal{G}_{\text{prob}} = (V_{\text{prob}}, E_{\text{prob}})$ where each vertex $v_i \in V_{\text{prob}}$ contains the parameters of the joint probability distribution $P(t_i, \mathbf{x}_i \mid o_i)$ and each edge $e_i \in E_{\text{prob}}$ contains the existence probability of a support relation $P(e_i \in \mathtt{SUPP} \mid \mathcal{O})$. Figure 2 shows an example of a probabilistic support graph.

2.2 Extraction from Point Clouds

This subsection describes how the probabilistic scene and support relation representation can be extracted from an input point cloud of a scene. The extraction algorithm consists of four steps: segmentation, geometry estimation, sampling, and support relations analysis (see Figure 1).

First, we segment the input point cloud using state-of-the-art algorithms [9]. For each segment in the point cloud, we estimate the joint probability distribution $P(t_i, \mathbf{x}_i \mid o_i)$ via a modified RANSAC algorithm. During the geometric primitive fitting using RANSAC, points are randomly drawn from the segment. For each primitive type, the state vector is estimated based on the drawn points.

Then, the number of inlier points and the sum of squared errors to all points in the segment is computed. Instead of selecting the primitive parameters which produced the maximum number of inliers, we collect all fitted primitive types $t_{i,j}$, state vectors $\mathbf{x}_{i,j}$, and the corresponding sum of squared errors $\varepsilon_{i,j}$ where j indicates the RANSAC iteration count. Given the total number of RANSAC iterations k, we can now estimate the discrete probability distribution over primitive types $P(t_i \mid o_i)$ by counting the occurrences of each concrete primitive type T:

$$P(t_i = T \mid o_i) = \frac{\sum_{j=1}^{k} (t_{i,j} = T)}{k}$$

The error can be used to calculate the weighted mean and covariance for the state vectors per primitive type. For each object o_i and primitive type t_i in each RANSAC iteration j, we assign a normalized weight $w_{i,j}$ using a Boltzmann distribution over the error $\varepsilon_{i,j}$. The constant β is a parameter, which controls the impact of errors on the weight.

$$w_{i,j} = \frac{e^{-\varepsilon_{i,j}/\beta}}{\eta}, \quad \eta = \sum_{j=1}^{k} e^{-\varepsilon_{i,j}/\beta}$$

To sample a set of concrete primitives from this geometric distribution, we sample a primitive per object o_i. First, the primitive type t_i^s is sampled from the discrete distribution $t_i^s \leftarrow P(t_i \mid o_i)$. Then, the state vector $\mathbf{x}_i^s$ can be drawn from the dependent multi-variate Gaussian distribution $\mathbf{x}_i^s \leftarrow P(\mathbf{x}_i \mid t_i = t_i^s, o_i)$. This results in a set of sampled primitives $\mathcal{P}_s$:

$$\mathcal{P}_s = \{p_1^s, p_2^s, \ldots, p_n^s\}, \quad p_i^s = (t_i^s, \mathbf{x}_i^s)$$

In the last step, we want to compute the existence probabilities of support relations given a geometric scene distribution $P(\mathcal{P} \mid \mathcal{O})$. We approximate the existence probabilities via Monte Carlo simulation. First, we generate m sets of primitives $\{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_m\}$ by sampling from the geometric scene distributions. Then, we extract pairwise support relations between the sampled primitives using static equilibrium and support polygon analysis presented in [4]. This way, we can estimate the existence probability for a support relation between two objects o_i and o_j by counting the occurrences of concrete support:

$$P\left(\mathtt{SUPP}(o_i, o_j) \mid \mathcal{O}\right) \approx \frac{1}{m} \cdot \sum_{k=1}^{m} \mathtt{SUPP}(p_i^k, p_j^k)$$

where $p_i^k, p_j^k \in \mathcal{P}_k$ are primitives sampled from their objects' geometric probability distributions. As $m \to \infty$, this estimate converges to the true value.

$$\lim_{m \to \infty} \left(\frac{1}{m} \cdot \sum_{k=1}^{m} \mathtt{SUPP}(p_i^k, p_j^k) \right) = P(\mathtt{SUPP}(o_i, o_j) \mid \mathcal{O})$$

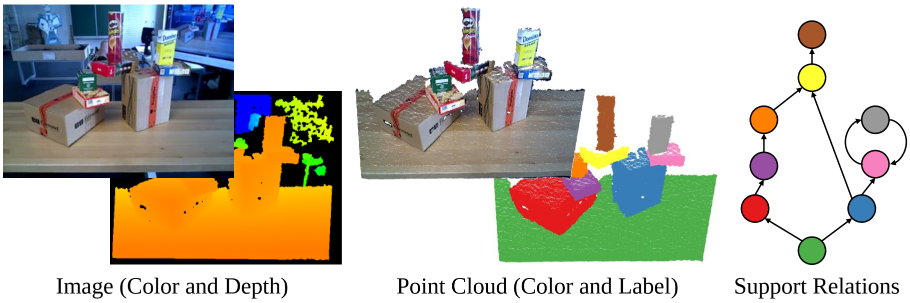

Fig. 3. A single entry in the KIT Support Relation (KIT-SR) dataset consists of color and depth images of the scene, a colored and labeled point cloud as well as manually annotated support relations.

3 Experiments

We present experimental results for the extraction of support relations from point cloud data. First, we introduce the two datasets and the evaluation procedure. Then, we present and discuss the results. The code and a reproduction guide for the results presented in this section are available online[2].

3.1 Evaluation Datasets

We present the KIT Support Relation (KIT-SR) dataset, which contains table-top scenes with a varying number of objects and support relations. We provide full pixel-wise segmentation and support relation annotations for each of the 60 scenes. The dataset is available online[3]. Figure 3 shows a scene from the dataset and the corresponding available data.

Additionally, we use the Object Segmentation Dataset (OSD, used in [7]) to compare our probabilistic support relation extraction with the deterministic approaches presented in [4,5]. The OSD contains table-top scenarios with varying object geometry and scene complexity, ranging from simple scenes with two objects to cluttered scenes with up to 16 objects. Since the dataset focuses on object segmentation, we added ground truth annotations for support relations to all 111 scenes by hand. These annotations are also available online.

3.2 Evaluation Procedure

The evaluation compares our novel probabilistic support extraction (Prob. Supp. Extr.) with two deterministic approaches. The first deterministic approach is based on static equilibrium analysis (St. Eq., [5]) and serves as a baseline for the comparison. The second deterministic approach extends the static equilibrium

[2] https://gitlab.com/h2t/interactive-scene-exploration/-/wikis/ISER-2020

[3] https://gitlab.com/h2t/interactive-scene-exploration/-/wikis/KIT-SR

analysis with support polygon analysis (St. Eq. + Supp. Poly., [4]) to detect top-down support relations, i. e. an object is supported by an object that is geometrically above it.

We evaluate the three approaches on the datasets by computing the metrics precision, recall, F1 score, and Brier score. Given the ground truth support relation $\texttt{SUPP}_{\text{GT}}$ and a hypothesis $\texttt{SUPP}_{\text{Hyp}}$ generated by one of the extraction methods, we can calculate precision and recall as follows:

$$\text{Precision} = \frac{|\texttt{SUPP}_{\text{GT}} \cap \texttt{SUPP}_{\text{Hyp}}|}{|\texttt{SUPP}_{\text{Hyp}}|}, \quad \text{Recall} = \frac{|\texttt{SUPP}_{\text{GT}} \cap \texttt{SUPP}_{\text{Hyp}}|}{|\texttt{SUPP}_{\text{GT}}|}.$$

The F1 score is the harmonic mean of precision and recall:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

When working with binary classification metrics and a probabilistic support representation, we first need to binarize the probabilities, i. e. decide whether support exists or not. We define support between $\texttt{A}$ and $\texttt{B}$ to exist if the existence probability is above the threshold 0.5:

$$P(\texttt{SUPP}(\texttt{A}, \texttt{B})) > 0.5$$

As a proper score function, the Brier score can be directly used with probabilities. It measures the accuracy of our probabilistic support relation extraction. The Brier score gets smaller when the hypothesis gets more accurate compared to the ground truth. The probabilities $P_{\text{Det}}(\texttt{SUPP}(\texttt{A}, \texttt{B}))$ for the deterministic approaches are set to 1 if support exists between objects $\texttt{A}$ and $\texttt{B}$, 0 otherwise.

$$\text{Brier Score} = \frac{1}{|\mathcal{O} \times \mathcal{O}|} \sum_{\texttt{A},\texttt{B} \in \mathcal{O} \times \mathcal{O}} (P_{\text{Hyp}}(\texttt{SUPP}(\texttt{A}, \texttt{B})) - P_{\text{GT}}(\texttt{SUPP}(\texttt{A}, \texttt{B})))^2$$

3.3 Results

Figure 4 and table 1 show the evaluation results on the KIT-SR dataset and the OSD. We can see a significant improvement in precision and a moderate increase in recall when comparing the probabilistic with the deterministic approaches. The Brier score is also significantly improved. Furthermore, the addition of support polygon analysis [4] yields better results than the static equilibrium analysis [5] but remains behind the probabilistic approach. The false positives and false negatives of the deterministic approaches are mostly caused by uncertainties about object geometry due to the dataset only containing single viewpoint clouds. The probabilistic approach captures these uncertainties and avoids making hard deterministic decisions early.

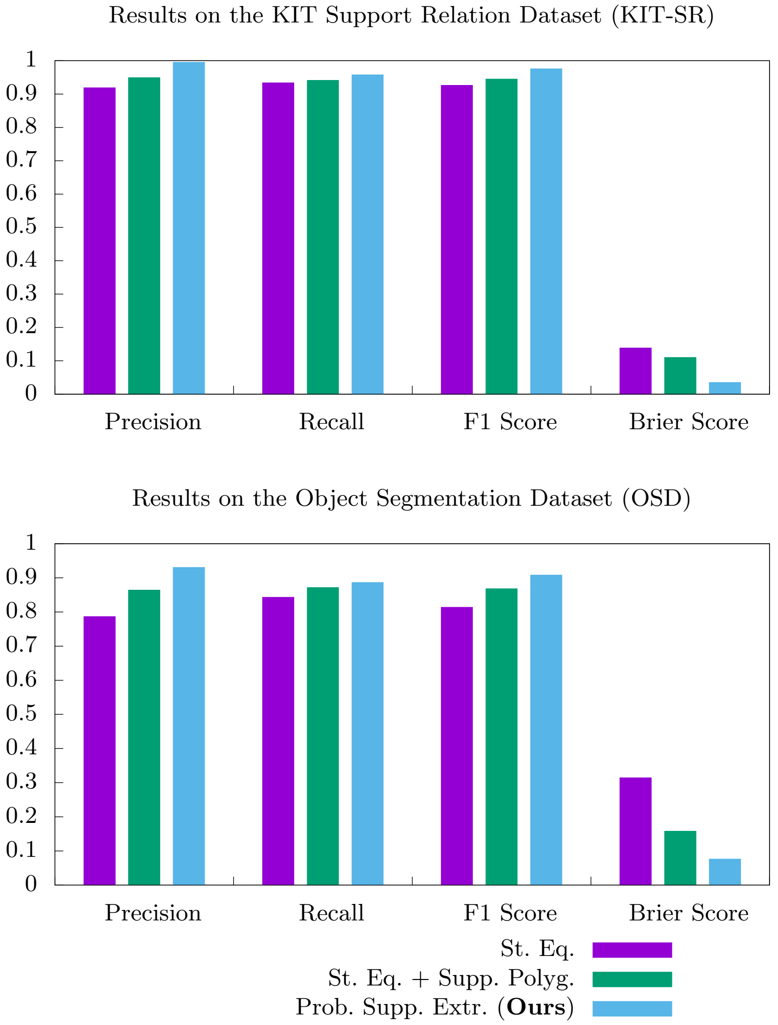

Fig. 4. Evaluation of the three approaches using the evaluation metrics precision, recall, F1 score, and Brier score on the two datasets KIT-SR and OSD. Higher values for precision, recall, and F1 score indicate better results, while lower values for the Brier score indicate better results.

Table 1. Evaluation results on the two datasets.

Dataset	**Method**	**Precision**	**Recall**	**F1**	**Brier Score**
		(higher is better)			(lower is better)
KIT-SR	St. Eq. [5]	0.919	0.934	0.926	0.138
	St. Eq. + Supp. Poly. [4]	0.949	0.941	0.945	0.109
	Prob. Supp. Extr. (**Ours**)	0.994	0.957	0.976	0.034
OSD	St. Eq. [5]	0.786	0.842	0.813	0.314
	St. Eq. + Supp. Poly. [4]	0.864	0.872	0.868	0.158
	Prob. Supp. Extr. (**Ours**)	0.931	0.886	0.908	0.076

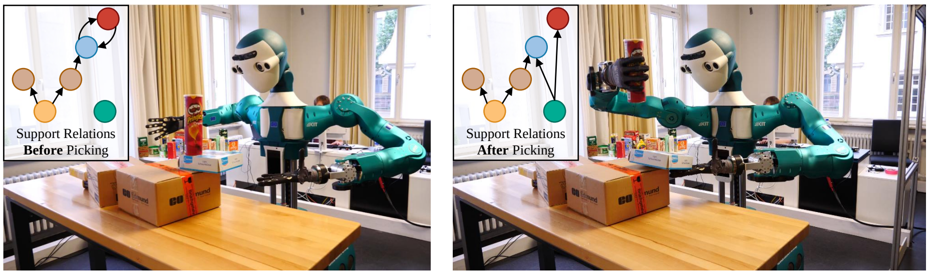

Fig. 5. ARMAR-6 picks a chips' can with the right hand and supports another object with the left hand to prevent the scene from collapsing.

3.4 Robot Experiments

Fig. 5 shows such a scenario, where the humanoid robot ARMAR-6 tries to pick the chips can from a stack of objects. Using the probabilistic support relations, the robot is able to infer that the box below the chips can is likely to fall, and prevents this by supporting the box with the left hand. A video of the experiment is available online[4].

4 Conclusion

In this paper, we have shown that a probabilistic representation of perceived object geometry and their support relations improves the support detection rate significantly compared to existing deterministic approaches. Occlusions and partial views on objects cause problems for the deterministic support analysis, which can be overcome by the proposed probabilistic approach. This novel representation allows a robot to reason about the safety and effects of its actions. In future work, we plan to combine this probabilistic scene representation with action effect prediction to support action sequence planning.

[4] `https://youtu.be/VB0vr5w7KhA`

Acknowledgment The research leading to these results has received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 146371743 – TRR 89 Invasive Computing and from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No. 731761 (IMAGINE).

References

1. Asfour, T., Wächter, M., Kaul, L., Rader, S., Weiner, P., Ottenhaus, S., Grimm, R., Zhou, Y., Grotz, M., Paus, F.: Armar-6: A high-performance humanoid for human-robot collaboration in real world scenarios. IEEE Robotics & Automation Magazine **26**(4), 108–121 (2019). https://doi.org/10.1109/MRA.2019.2941246
2. Battaglia, P.W., Hamrick, J.B., Tenenbaum, J.B.: Simulation as an engine of physical scene understanding. Proceedings of the National Academy of Sciences **110**(45), 18327–18332 (2013). https://doi.org/10.1073/pnas.1306572110
3. Jia, Z., Gallagher, A.C., Saxena, A., Chen, T.: 3d reasoning from blocks to stability. IEEE transactions on pattern analysis and machine intelligence **37**(5), 905–918 (2014)
4. Kartmann, R., Paus, F., Grotz, M., Asfour, T.: Extraction of physically plausible support relations to predict and validate manipulation action effects. IEEE Robotics and Automation Letters (RA-L) **3**(4), 3991–3998 (2018)
5. Mojtahedzadeh, R., Bouguerra, A., Schaffernicht, E., Lilienthal, A.J.: Support relation analysis and decision making for safe robotic manipulation tasks. Robotics and Autonomus Systems **71**, 99–117 (2015). https://doi.org/10.1016/j.robot.2014.12.014
6. Panda, S., Hafez, A.A., Jawahar, C.: Single and multiple view support order prediction in clutter for manipulation. Journal of Intelligent & Robotic Systems **83**(2), 179–203 (2016)
7. Richtsfeld, A., Mörwald, T., Prankl, J., Zillich, M., Vincze, M.: Segmentation of unknown objects in indoor environments. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 4791–4796. IEEE (2012)
8. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: European conference on computer vision. pp. 746–760. Springer (2012)
9. Stein, S.C., Schoeler, M., Papon, J., Worgotter, F.: Object partitioning using local convexity. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 304–311 (2014)
10. Xue, F., Xu, S., He, C., Wang, M., Hong, R.: Towards efficient support relation extraction from rgbd images. Information Sciences **320**, 320–332 (2015)
11. Yang, C., Lan, X., Zhang, H., Zhou, X., Zheng, N.: Visual manipulation relationship detection with fully connected crfs for autonomous robotic grasp. In: 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO). pp. 393–400. IEEE (2018)
12. Zhang, H., Lan, X., Zhou, X., Tian, Z., Zhang, Y., Zheng, N.: Visual manipulation relationship network for autonomous robotics. In: 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids). pp. 118–125. IEEE (2018)
13. Zheng, B., Zhao, Y., Yu, J., Ikeuchi, K., Zhu, S.C.: Scene understanding by reasoning stability and safety. International Journal of Computer Vision **112**(2), 221–238 (2015)