SPONGE: Sequence Planning with Deformable-ON-Rigid Contact Prediction from Geometric Features

Tran Nguyen Le[1], Fares J. Abu-Dakka[2], Ville Kyrki[1]

I. INTRODUCTION

Volumetric deformable objects, present in various forms, are prevalent in many aspects of daily life, including food, toys, or internal organs of humans. Successful manipulation of such objects can lead to numerous practical applications in areas such as surgical manipulation, or food processing where robots can be used to make pizza dough [1], cut fruits [2], [3], or in healthcare where robots can be used to rearrange objects in target configurations [4], assistive dressing [5]–[7], clean dishes [7]. These tasks are trivial for humans because not only we possess remarkable dexterity but we also excel at task planning, as demonstrated by our ability to perceive objects at hand, and develop a plan to complete the task with precision and accuracy in less than a second.

Planning manipulation tasks involving interactions between deformable and rigid objects, such as wiping a curved surface with a deformable tool, is difficult due to the challenge in predicting such interactions. Majority existing works disregard the interaction between the deformable tool and target objects, and focus only on the control aspect of the tasks [8]. Only recently have some researchers started to investigate how to estimate and harness such interactions in different tasks such as assistive dressing [5]–[7], or food processing [1]. In the literature, researchers studied the interaction between complex deformable objects such as human hands [9], [10] or cloth-like objects [5]–[7] and rigid bodies by looking at the concept of *contact reasoning* where the location of contact and the magnitude of applied forces are estimated once the two bodies interact with each other. However, this *contact reasoning* concept is more suitable for the control aspect than for the planning aspect due to its ability to track the interaction in real time. Thus, the question of how to predict the interaction between volumetric deformable tool and rigid objects and exploit such interactions for planning remains open.

To address the aforementioned open issues, we propose **S**equence **P**lanning with deformable-**ON**-rigid contact prediction from **GE**ometric features (SPONGE), a sequence planning pipeline powered by a contact prediction model that predicts contact between deformable and rigid bodies, with the aim of providing robots with the aforementioned human-like planning skill in order to efficiently automate downstream deformable object manipulation tasks such as cleaning dishes (Fig. 1). Instead of *contact reasoning*, in this paper we tackle the concept of *contact prediction* of a 3D deformable tool acting on rigid objects, which is better suited for planning purposes. We take a data-driven approach with physics-based simulation to model the interactions between 3D deformable objects and rigid objects. We then use PointNet [11] architecture to form a mapping from point-cloud observation of the target object, and pose of the deformable tool to 3D representation of the contact points between the two bodies. The trained contact prediction model is then used as the driving force behind the planning of a subsequent task. Finally, we deploy SPONGE in the real world to demonstrate that the contact prediction model trained only with synthetic data from physics-based simulation can help to produce an efficient plan for a manipulation task to be successfully executed in the real world.

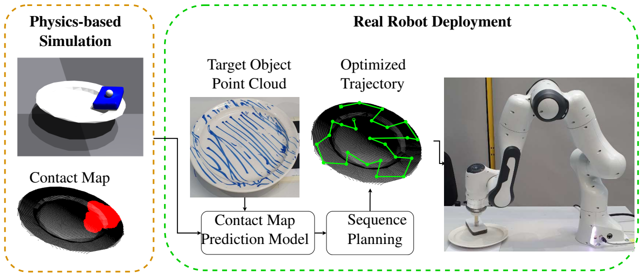

Fig. 1: SPONGE deployed in the real world to accomplish a dish cleaning task with a deformable sponge. Given a point cloud of the target objects, SPONGE powered by a contact map prediction model trained in simulation, plans an optimal trajectory aiming at achieving full area coverage of target objects with the least number of waypoints.

II. METHOD

The proposed planning pipeline shown in Fig. 2 consists of two important steps: (i) learning to predict the contact between the deformable tool and target objects, (ii) planning an area-coverage trajectory on the target objects.

A. Contact Map Prediction Model

Knowledge of the contact area between two bodies is crucial when it comes to planning manipulation tasks associated with the interaction between two bodies [12]. Let us consider Fig. 3, in which a human manipulates a deformable sponge to clean rigid dishes. From the figure, we can see that the contact areas between the sponge and the dishes are highly dependent on the contact location and the geometric features

This work was supported by CHIST-ERA project IPALM (326304). Abu-Dakka is supported by the European project euROBIN under grant agreement No 101070596. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

[1] Intelligent Robotics Group at the Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, Finland. {firstname.lastname}@aalto.fi

[2] MIRMI, Technische Universität München, 80992 München, Germany. Part of the research presented in this work was conducted when F. Abu-Dakka was at Aalto University. fares.abu-dakka@tum.de

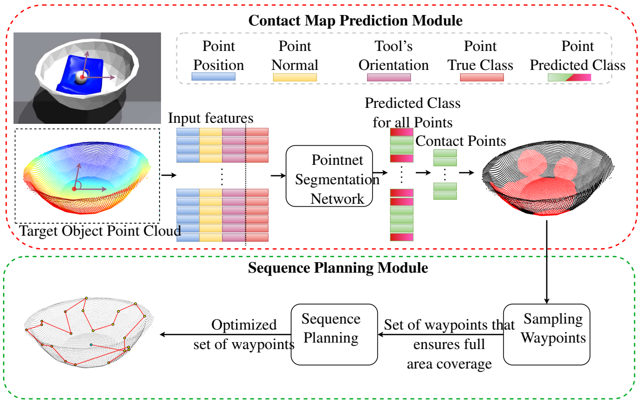

Fig. 2: The proposed planning pipeline consists of: a contact map prediction module learns from target object point clouds, which form the input to a dense point cloud network to produce per-point contacts; a sequence planning module that harnesses the trained prediction model to generate an optimal trajectory to accomplish the task.

Fig. 3: Different tactics of human to create contact with various complex surface profiles using a deformable sponge.

at that contact location. For example, in the case of objects with curved surfaces (Fig. 3 a,b,c), the deformability of the sponge allows it to conform to the curved surface to cover more area of the object. Let us examine Fig. 3 a, one can wipe both the bottom and the wall of the pan simply by pressing the sponge on the intersection line, which would not be possible if the sponge was rigid. Inspired by this behavior, we want to develop a model that learns a mapping from the geometric features of the rigid objects to the contact areas between the deformable and the rigid objects.

To this end, we propose using a dense point cloud network to model the contact information between the deformable tool and the target object. Specifically, we use a Pointnet segmentation network [11], which, given an input point cloud of the target object $\mathcal{P}_O$ produces per-point outputs. It is worth noting that the Pointnet segmentation network is not trained to do segmentation, as the name implies, but to predict per-point contact class, which indicates whether a point of the target object is in contact with the deformable tool or not.

The input of the network includes the position and normal vector of each point $p_i \in \mathcal{P}_O$ and a feature vector associated with each point p_i. Point positions are normalized to the zero mean, enabling the model to be invariant for point-cloud translations. The feature vector is a two-dimensional vector $[sin(\theta), cos(\theta)]$ representing the orientation of the deformable tool around the Z axis at the contact location p_c. For points that are not the contact location, the feature vectors are defined as [0,0]. The output of the contact prediction model is the contact class of each point p_i of the target object, where 1 is in contact and 0 is not in contact with the deformable tool. The proposed network is trained with supervised learning manner on a synthetic dataset with the Binary Classification Loss function. Readers are referred to [13] for more details related to the synthetic training dataset.

B. Area Coverage Planning Under Deformations

We address the problem of area coverage planning under deformations, where our goal is to plan an optimal trajectory that covers the entire surface of the target objects using a deformable tool while harnessing the learned contact map prediction model. The proposed algorithm consists of two steps: 1) sampling waypoints, where we solve the Set Cover problem [14] to sample sets of waypoints that ensure 100% of the deformable tool's area coverage of the target objects, 2) sequence planning, where we choose and optimize the optimal trajectory from the obtained sets of waypoints.

In the first step, the planning algorithm takes the point cloud of the target object along with the number of sets to be sampled and produces $\mathcal{T}$ containing n_{sets} sets of contact points on the surface of the target object. We achieve this by solving the set cover problem using a heuristic bottom-up sampling algorithm, where we first randomly sample a contact point on the target object surface, predict the contact areas at that point, and remove all the points that are in contact from the target object point cloud. This process is repeated until the remaining point cloud of the target object is empty, indicating that we have covered the entire object.

Once $\mathcal{T}$ is obtained, we proceed to the sequence planning step, where the goal is to produce an optimal trajectory that achieves full area coverage of the target objects while minimizing a certain cost measure, such as the travel distance. We frame this problem in relation to the well-known travel salesman problem (TSP) [15]. We achieve this by solving the TSP with the 2-Opt algorithm [16] for the best set of waypoints $\mathcal{T}_j \in \mathcal{T}$, which ensures 100% of the area coverage of the target objects with the least number of waypoints. More formally, the optimal trajectory is defined as

$$\mathcal{T}_{opt} = TSP(\underset{\mathcal{T}_j \in \mathcal{T}}{\arg\min}\, len(\mathcal{T}_j)) \tag{1}$$

III. Experiments and Results

We evaluate each component of our *SPONGE* pipeline: contact map prediction and sequence planning in both simulation and real world.

A. Contact Map Prediction Result

We access the performance of the proposed prediction model based on the contact prediction F1 score on the test dataset, which is approximately **0.95**. This high score indicates that the proposed model is capable of accurately predicting the contact map, given only the applied position and the rotation of the sponge. Fig. 4 qualitatively compares the predicted contact map from the proposed model with the ground truth in the simulation. As shown, the predicted

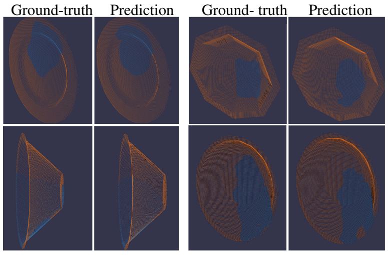

Fig. 4: Visualization of the contact map predictions on the test dataset and the ground-truth in simulation. Left columns show the ground-truth and the right columns show the prediction. Blue indicates points that are in contact with the deformable tool, while orange denotes points that are not in contact with the deformable tool.

TABLE I: The average area coverage (%) and average number of waypoints over 20 trajectories on the target objects in simulation. ↑: higher the better

Object ID	Area Coverage	Number of Waypoints
1	89.5	19.3
2	97.3	13.3
3	100	18.1
4	96.2	14.1
5	87.6	23.6
6	97.9	22.3
7	98.7	12.4
8	88.5	32.7
9	92.1	22.4
10	94.6	20.2
All ↑	94.27	-

contact maps are qualitatively similar to ground truth. It should be noted that the model was able to capture the correlation between the geometric features of the target object and the contact map between the two bodies. For example, let us examine the bowls shown in Fig. 4, because of the deformability of the tool, it conforms to the curvature of the bowls when applied to the side of the bowl. Our proposed model was able to capture this behavior by taking into account the pose of deformable tool, and the local features close to the applied location.

B. Planning in Simulation Result

We investigate the quality of the generated trajectories in the context of the dish cleaning task in the Isaac Gym simulator using the same environment. The position and contact forces of the sponge are recorded during the execution process. As the objective of the task is to cover the entire surface of the target objects with the planned trajectory, we quantify the quality of the trajectory by the area coverage (*i.e.,* proportion of the contact points to the total population of point clouds). For each object, we randomly initialized its starting position 20 times and evaluated the best trajectory. In total, we evaluated 200 trajectories on all ten objects.

Table I presents the average area coverage over 20 optimal trajectories on all objects. As expected, the number of waypoints needed to cover the entire surface increases as the size of the target objects increases. These results clearly show that the proposed planning pipeline is capable of producing high-quality trajectories that cover approximately 94% the surfaces of different objects with varying geometry and curvatures.

C. Real Robot Deployment

To investigate how well the proposed pipeline performs in the real world, we conducted an experiment that performs a dish cleaning task with a Franka Emika Panda equipped with a hemispherical finger attached to a deformable sponge, as shown in Fig. 1. The dimensions and material characteristics of the sponge used in the real-world experiment are identical to those of the sponge in the simulation. The goal is to clean all the blue marker writings, which represent dirt on the top surface of two target objects used in this experiment are a bowl and a plate that are not from any dataset.

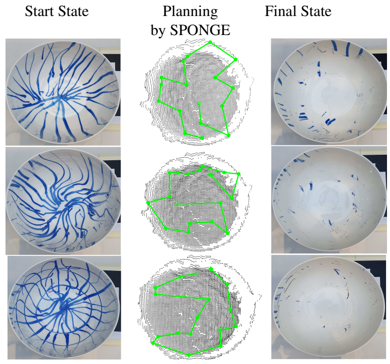

Fig. 5: Qualitative visualizations of *SPONGE* for real-world dish cleaning task. Columns headed by Start State are the target object with blue marker writings denoting dirt need to be removed. The optimal coverage trajectories planned by *SPONGE* (solid green lines) are shown in columns headed by Planning by SPONGE. The columns headed by Final State are when robot is done executing the early planned trajectories.

TABLE II: The average area coverage (%) and average number of waypoints over 5 trajectories on the target objects. ↑: higher the better

Object	Area Coverage	Number of Waypoints
Plate	94.5	20.4
Bowl	96.2	17.4
All ↑	95.35	-

Fig. 5 shows qualitative results of the real robot deployment of *SPONGE* on the two target objects. We can see that the robot has successfully accomplished the dish cleaning task by removing **almost** all of the blue marker writings from the surfaces of the target objects with less than 20 waypoints. This observation was further reflected by the quantitative results shown in Table II with an impressive area coverage of more than **95%** in two different geometries.

D. Discussion & Limitation

Several limitations of *SPONGE* can be addressed as follow. First, the current contact map prediction model lacks real-time knowledge of the en route contact map while moving from one contact point to another. This additional knowledge can potentially increase the efficiency of the planning pipeline, so that fewer waypoints would be needed to cover the entire surface of the objects.

Second, in the sequence planning module, since we just randomly sample contact points until the entire surface is covered, the resulting trajectories may look counter-intuitive compared to smooth spiral-shaped trajectories of humans.

Finally, the planned trajectories are executed in an open-loop manner, where we omit the actual contact happening between the two objects during the manipulation. This information is important in reacting and adapting trajectories to uncertainties such as incorrect contact map prediction, or displacement of the target object during execution procedure.

References

[1] X. Lin, C. Qi, Y. Zhang, Z. Huang, K. Fragkiadaki, Y. Li, C. Gan, and D. Held, "Planning with spatial-temporal abstraction from point clouds for deformable object manipulation," in *6th Annual Conference on Robot Learning*, 2022.

[2] E. Heiden, M. Macklin, Y. S. Narang, D. Fox, A. Garg, and F. Ramos, "DiSECt: A Differentiable Simulation Engine for Autonomous Robotic Cutting," in *Proceedings of Robotics: Science and Systems*, Virtual, July 2021.

[3] Z. Xu, Z. Xian, X. Lin, C. Chi, Z. Huang, C. Gan, and S. Song, "Roboninja: Learning an adaptive cutting policy for multi-material objects," in *arXiv*, 2023.

[4] B. Shen, Z. Jiang, C. Choy, L. J. Guibas, S. Savarese, A. Anandkumar, and Y. Zhu, "Acid: Action-conditional implicit visual dynamics for deformable object manipulation," *Robotics: Science and Systems (RSS)*, 2022.

[5] Z. Erickson, A. Clegg, W. Yu, G. Turk, C. K. Liu, and C. C. Kemp, "What does the person feel? learning to infer applied forces during robot-assisted dressing," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 6058–6065.

[6] Z. Erickson, H. M. Clever, G. Turk, C. K. Liu, and C. C. Kemp, "Deep haptic model predictive control for robot-assisted dressing," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 4437–4444.

[7] Y. Wang, D. Held, and Z. Erickson, "Visual haptic reasoning: Estimating contact forces by observing deformable object interactions," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 11 426–11 433, 2022.

[8] J. Kim, A. K. Mishra, R. Limosani, M. Scafuro, N. Cauli, J. Santos-Victor, B. Mazzolai, and F. Cavallo, "Control strategies for cleaning robots in domestic applications: A comprehensive review," *International Journal of Advanced Robotic Systems*, vol. 16, no. 4, p. 1729881419857432, 2019.

[9] S. Brahmbhatt, C. Tang, C. D. Twigg, C. C. Kemp, and J. Hays, "ContactPose: A dataset of grasps with object contact and hand pose," in *The European Conference on Computer Vision (ECCV)*, August 2020.

[10] H. Jiang, S. Liu, J. Wang, and X. Wang, "Hand-object contact consistency reasoning for human grasps generation," in *Proceedings of the International Conference on Computer Vision*, 2021.

[11] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 77–85.

[12] Z. Weng, F. Paus, A. Varava, H. Yin, T. Asfour, and D. Kragic, "Graph-based task-specific prediction models for interactions between deformable and rigid objects," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 5741–5748.

[13] T. N. Le, F. J. Abu-Dakka, and V. Kyrki, "Sponge: Sequence planning with deformable-on-rigid contact prediction from geometric features," 2023.

[14] V. Chvatal, "A greedy heuristic for the set-covering problem," *Mathematics of Operations Research*, vol. 4, no. 3, pp. 233–235, 1979.

[15] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook, 2007.

[16] G. A. Croes, "A method for solving traveling-salesman problems," *Operations Research*, vol. 6, no. 6, pp. 791–812, 1958.