MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Estimation of Extrinsic Contact Patch for Stable Placement

Ota, Kei; Jha, Devesh K.; Jatavallabhula, Krishna; Kanezaki, Asako; Tenenbaum, Joshua B.

TR2024-018 March 12, 2024

Abstract

Precise perception of contact interactions is essential for the design of very precise and fine manipulation skills for robots. In this paper, we present the design of feedback skills for robots that must learn to stack complex-shaped objects on top of each other (see Fig. 1). To design such a system, a robot should be able to reason about the stability of placement from very gentle contact interactions. Our results demonstrate that it is possible to infer the stability of object placement based on tactile readings during contact formation between the object and its environment. In particular, we estimate the contact patch between a grasped object and its environment using force and tactile observations to estimate the stability of the object during a contact formation. The contact patch could be used to estimate the stability of the object up on release of grasp. The proposed method is demonstrated on various pairs of objects that are used in a very popular board game.

IEEE IROS 2023 Workshop on Visuo-Tactile Perception, Learning, Control for Manipulation and HRI

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Estimation of Extrinsic Contact Patch for Stable Placement

Kei Ota[1,2], Devesh K. Jha[3], Krishna Murthy Jatavallabhula[4], Asako Kanezaki[2], and Joshua B. Tenenbaum[4]

***Abstract*— Precise perception of contact interactions is essential for the design of very precise and fine manipulation skills for robots. In this paper, we present the design of feedback skills for robots that must learn to stack complex-shaped objects on top of each other (see Fig. 1). To design such a system, a robot should be able to reason about the stability of placement from very gentle contact interactions. Our results demonstrate that it is possible to infer the stability of object placement based on tactile readings during contact formation between the object and its environment. In particular, we estimate the contact patch between a grasped object and its environment using force and tactile observations to estimate the stability of the object during a contact formation. The contact patch could be used to estimate the stability of the object up on release of grasp. The proposed method is demonstrated on various pairs of objects that are used in a very popular board game.**

I. Introduction

Humans can perform very complex and precise manipulation tasks effortlessly. The human vision and tactile perception system is very advanced, and we can perform physical reasoning of different objects in various complex contact formations very efficiently. We can monitor and control complex contact formation in a closed-loop fashion during various manipulation tasks. Designing such reactive and closed-loop robotic systems remains elusive. Consider, for example, the stacking task shown in Fig. 1. To do this, a robot should have the ability to estimate the stability of the object as it is trying to place the grasped object on the bottom object so that it can release the object in a stable configuration. Estimating the stability from vision alone could be insufficient due to occlusions from other objects during placement. Thus, we present a closed-loop system that can reason about the stability of the object using tactile signals from extrinsic contact formation during placement.

We believe that object stability could be estimated from the contact forces experienced by an object during placement. The stability of an object is governed by the relative location of the environmental contact and the center of mass location of the object. The forces experienced by a force-torque sensor mounted on the wrist of the robot depend on the contact patch between the object and its environment, as well as the geometric and physical properties of the object. As a simplification, we assume that the geometry of the objects is fixed, so the robot works with known pieces. Under this assumption, we estimate the stability of the object during placement using tactile signals. In particular, we try to estimate the contact patch between the object and the environment using tactile signals. The robot can then attempt to maximize the contact patch to achieve a stable placement of the object. This is demonstrated using several pairs of objects from a popular board game where the objective is to incorporate a new block on an existing tower without destabilizing the tower.

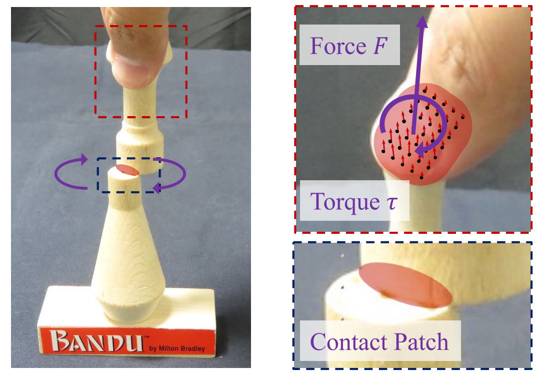

Fig. 1: In this work, we try to understand the local contact phenomena during placement of an object in an environment with partial support. Since stability of the object depends on the contact formation between the object and the environment, we propose a method to estimate contact patch given force as well as tactile observations during the contact phenomena. This is very similar to how a human would perform this task under partial observability.

II. Related Work

Robot stacking. Several studies have addressed the problem of robot stacking through various approaches. These include learning to schedule auxiliary tasks for reinforcement learning (RL) [1], combining demonstrations and RL [2], [3], employing sim-to-real transfer [2], [4], [5], and using task-and-motion planning [6]. The focus of these works primarily revolves around stacking simple cubes. Lee *et al.* [7] propose a benchmark that introduces relatively irregular rectangles generated by deforming cubes. However, these objects still maintain convexity and simplicity. Furrer *et al.* [8] and Yifang *et al.* [9] have explored the stacking of irregular stones. Nonetheless, these studies make assumptions about known geometries and assume that the stones possess wide support and high friction, simplifying the problem and enabling basic pick-and-place strategies. In contrast, our research considers the local contact phenomenon where the

[1]Kei Ota is with Mitsubishi Electric Corporation, Kanagawa, Japan. Ota.Kei@ds.MitsubishiElectric.co.jp

[2]Devesh K. Jha is with Mitsubishi Electric Research Labs, Cambridge, MA, USA.

[3]Kei Ota and Asako Kanezaki are with Tokyo Institute of Technology, Japan.

[4]Krishna Murthy Jatavallabhula and Joshua B. Tenenbaum are with Massachusetts Institute of Technology, Cambridge, MA, USA.

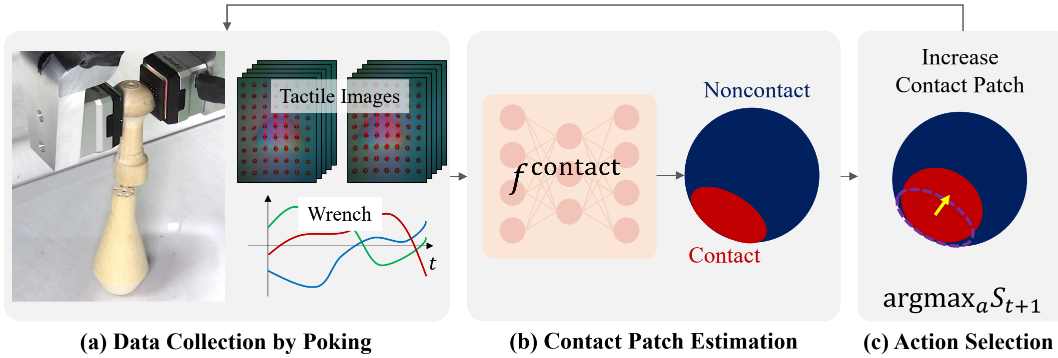

Fig. 2: Our proposed method comprises of three components. First, a robot uses a probing action to establish contact between the grasped and its environment. During probing, it acquires a sequence of force/torque measurements and tactile images. Based on the collected data, we estimate the contact patch between the grasped object and the bottom object. Subsequently, we estimate stability from the estimated contact patch. The robot can then maximize the contact patch during the probing attempt and release the object based on estimate of stability of the object up on release.

object can topple and fall, if not placed with proper support. Moreover, we remove assumptions regarding geometries of underlying objects, necessitating the estimation of stability through interactions.

External contact localization Prior works represent contacts as a set of points [10], [11] and lines [12], [13]. Although line contacts give us more information compared to point contacts, they require active exploration involving changes in the orientation of the gripper [12], [13], making it difficult to apply them in our setting where the tower is very unstable. The closest work to ours is the neural contact fields (NCF) of Higuera et al. [14], where the authors estimate the contact patch between a grasped object and its environment. While NCF is evaluated on a simulation and a limited number of objects, we tested our method on unknown geometries of the environment which can be used for an appropriate downstream task.

III. Problem Statement

We consider the problem of estimating the stability of a grasped object when in contact with its environment, in an attempt to release and place the object in a stable pose during a task. This happens to be a partially observable task as we can not observe the full state of the system, and thus stability needs to be estimated from sensor observations. We assume that the robot has access to tactile sensors co-located at the gripper fingers as well as a Force/Torque (F/T) sensor at the wrist of the robot. A certain contact formation is stable if the object can stay stable after release from the grasp.

The stability of a contact formation depends on the relative position of the center of mass of the object and the contact patch between the object and the environment. However, this can not be directly observed during a contact formation, and thus leads to partial-observability. A robot usually can observe force-torque signals and/or tactile images during the interaction. The observed signals depend not only on the contact formation but also on the geometry and physical parameters of the grasped object. Thus, although these data have a lot of information, these are all entangled and thus it is very difficult to extract specific information, e.g. estimate contact patch. To simplify the estimation problem, we make the following assumptions to limit the scope of current study:

1) Geometry and physical parameters of the grasped objects are fixed.
2) All objects are rigid and have flat surfaces.

It is important to emphasize that the robot is unfamiliar with the shape of the underlying objects and needs to explore a stable configuration through several probing attempts.

IV. Method

The main idea here is to estimate the contact patch between an object and its environment using force and tactile measurements. This is based on the fact that sensor observations are generated by the contact formation between the object and its environment. We propose a method consisting of three key parts. First, the robot estimates the contact patch between the grasped object and the bottom object. Then it can assess the stability from the estimated contact patch. Finally, the robot selects an action based on the estimated contact patch and stability; and releases the grasped object if the current configuration is stable; otherwise, moves the object towards a position that can improve stability. In this section, we describe more details on these three modules.

A. Contact Patch Estimation

As described in Sec. III, we explicitly estimate the extrinsic contact patch from a sequence of tactile images and force-torque measurements. During a duration of T seconds, the robot applies a downward force along the negative Z axis for d mm, while collecting data $s^{\text{Tac}}_{i,0:T}, s^{\text{FT}}_{i,0:T}$ from tactile and force-torque sensors at a frequency of 10 Hz with i denoting an index of data. We use a suitable impedance control to prevent the object from falling down by using excessive force. Specifically, $s^{\text{Tac}}_{i,0:T} = \{s^{\text{Tac}}_{i,t}\}^{T}_{t=0}$, where $s^{\text{Tac}}_{i,t} \in \mathbb{R}^{63}$ and $63 = 7 \times 9$ are the number of markers in column and row (see Fig. 2), which can be obtained by post-processing the

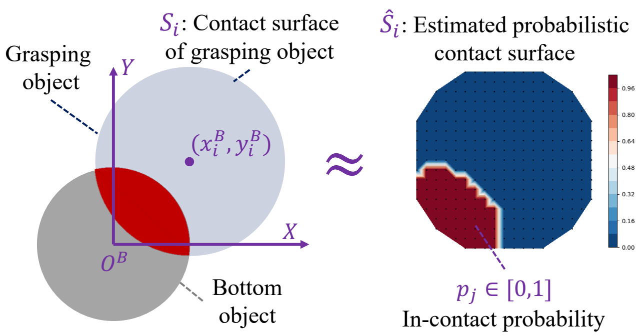

Fig. 3: Definition of the probabilistic contact surface. The left coordinate defines the displacements when data collection. The displacements (x_i^B, y_i^B) are manually added from the origin of the bottom object O^B. This displacement and ground truth contact surfaces of the two objects give the ground-truth contact surface S_i. We try to estimate the contact patch $\hat{S}_i$ using a neural network, which consists of a set of points p_j that represents the probability of contact or uncontacted.

Fig. 4: The Bandu pieces employed in our experiments consist of five distinct shapes. The first two pieces on the left serve as the bottom objects (or the environment), while the subsequent three on the right are designated as the (top) objects. These pieces have been assigned the following names, proceeding from left to right: *Short*, *Long*, *Mushroom*, *Barrel*, and *Pot*.

tactile image $I_{i,t}^{\text{Tac}}$. In addition, $s_{i,0:T}^{\text{FT}} = \{s_{i,t}^{\text{FT}}\}_{t=0}^{T}$, $s_{i,t}^{\text{FT}} \in \mathbb{R}^6$ are the force-torque measurements.

In the data collection process, we add displacements in the XY plane such as $x_i^B \sim \{x_{\texttt{min}}, x_{\texttt{max}}\}$ and $y_i^B \sim \{y_{\texttt{min}}, y_{\texttt{max}}\}$ whose origin is the center position of the contact surface of the lower object O^B (see Fig. 3), and the minimum and maximum positions are defined to ensure contact between the flat surfaces of the upper and lower objects. Since the top and bottom objects used for data collection are known in advance, it is possible to obtain a contact patch between the two surfaces based on the displacements x_i^B and y_i^B. To estimate the contact patch, we discretize the contact surface of the grasped object into N points $p_j, j \in \{1, \ldots, N\}$ each of these points corresponds to a specific location on the contact surface of the grasped object and represents the probability of being in contact or remaining uncontacted. Consequently, this process yields the representation of the estimated probabilistic contact surface denoted as $\hat{S}_i = \{p_1, \ldots, p_N\}$ between the two objects.

Finally, we train a model f^{contact} that takes the observed data $s_{i,0:T}^{\text{FT}}, s_{i,0:T}^{\text{Tac}}$ to generate the probability of contacts $\hat{S}_i$ by minimizing the binary cross entropy loss for each data point p_j.

B. Stability Estimation

Utilizing the estimated contact patch $\hat{S}_i$, denoted as a set of points $\{p_j\}$, we use it to assess the stability of the current configuration. This assessment process begins by constructing a convex hull $C_i \in \text{Conv}(\hat{S}_i)$ using points whose associated probability exceeds a predefined threshold denoted δ, thus effectively isolating reliable contact points. Subsequently, an evaluation is performed to ascertain whether this convex hull encompasses the position of the center of mass of the grasped object. In the affirmative case, the gripper releases the grasped object. Conversely, if this evaluation yields a negative result, the gripper moves towards a position to enhance stability by a policy described below.

C. Action Selection

We formulate a policy aimed at increasing the contact area in the subsequent step, guided by the estimated contact patch. This policy begins by calculating the central position of the convex hull, denoted as $\bar{p}$, and subsequently directs the robot to navigate in the negative direction relative to $\bar{p}$. Additionally, to mitigate a large movement at each step, we restrict movement within d^{move} mm if the norm exceeds d^{move}. We specifically set $d^{\text{move}} = 3$ [mm].

V. Experiments

A. Settings

Tactile sensor. We use a commercially available GelSight Mini [20] tactile sensor, which provides 320×240 compressed RGB images at a rate of approximately 25 Hz, with a field of view of 18.6 × 14.3 millimeters. We use gels that have 63 tracking markers so that the data from the tactile sensor are $x_t^{\text{Tac}} \in \mathbf{R}^{63}$.

Robot platform. The MELFA RV-5AS-D Assista robot, a collaborative robot with 6 DoF, is used in this study. The tactile sensor is mounted on the WSG-32 gripper (see Fig. 2). We use a Force-Torque (F/T) sensor, which is mounted on the wrist of the robot. The F/T sensor is used for two-fold: data collection for estimating the contact patch, and the stiffness control of the position-controlled robot.

Bandu We use pieces from *Bandu* for our experiment. Bandu is a toy game that involves stacking objects onto a base plate. The players take turns stacking these objects and compete to see who can stack the most. Each piece has a highly irregular shape, requiring robots to estimate stable placements based on the shape of the objects. Figure 4 illustrates the Bandu pieces used in this experiment. The challenge in the game is to accommodate an irregular piece into an existing tower without destabilizing the existing tower.

TABLE I: Comparison of the contact patch estimation performance on different input modalities measured by IoU (higher is better). Bold numbers show the best results among the three different input modalities. The *S* and *L* of the bottom object correspond to the *Short* and *Long* objects, respectively (see Fig. 4).

		Mushroom		Barrel		Pot		Mean
		S	L	S	L	S	L	
ID	FT	**76.2**	92.2	79.9	88.1	82.1	94.4	85.5
	Tac	66.6	82.4	76.8	79.7	78.5	86.8	78.5
	FT+Tac	75.7	**94.0**	**86.3**	**91.6**	**86.8**	**94.7**	**88.2**
OOD	FT	43.0	41.4	**36.4**	39.3	40.9	38.4	39.9
	Tac	**48.2**	36.6	33.9	**47.9**	37.0	**44.3**	41.3
	FT+Tac	43.6	**43.2**	34.9	41.0	**46.9**	39.1	**41.4**

B. Contact patch estimation

Settings As described in Sec. IV-A and illustrated in Fig. 4, we used three top pieces and two bottom pieces for data collection. The data acquisition process involved incremental movements of the top piece at intervals of 0.5 mm, originating from the center position and spanning a range of -10 mm to 10 mm in both the X and Y directions, as defined in Fig. 3. Consequently, the total dataset encompasses $41 \times 41 = 1681$ individual data points. Subsequently, for each pair of top and bottom objects, we trained the model f^{contact}. Performance evaluation of this model was performed using the intersection over union (IoU) metric. We tested the model with different input modalities, only force torque sensor, tactile sensors, and the combination of the two denoted as *FT*, *Tac*, and *FT+Tac*, respectively. The evaluation is carried out on the same and different bottom object, which we denote as *ID* (in-distribution) and *OOD* (out-of-distribution).

Results and analyses The results are presented in Table I. When comparing the various modalities, it is evident that the combination of tactile sensors and the force-torque sensor yields the most favorable performance. Consequently, for our subsequent experiments, we will utilize both of these modalities. However, it should be noted that the model performance on different bottom objects (OOD) generally exhibits a marked decline in comparison to its performance on the same bottom object (ID).

C. Stability Analysis

Settings Next, we assess whether the proposed framework predicts the stability from the estimated contact patch with the same data used in the previous experiment. The evaluation metric to assess stability is binary accuracy. Specifically, a positive outcome is recorded when both the model prediction and the ground truth yield concordant results, indicating either stability or instability.

Results and analyses Table II shows the qualitative results of the stability estimation. The results indicate that extrinsic contact patch detection would be useful in estimating stability when stacking objects. However, the performance is still not reliable, so we will add probabilistic inference to aggregate information from multiple touches to improve the performance of the stability estimation in the future.

TABLE II: Stability estimation performance measured by binary accuracy (higher is better).

	Mushroom		Barrel		Pot		Mean
	S	L	S	L	S	L	
ID	79.2	90.8	89.1	90.2	88.4	93.8	88.2
OOD	63.5	67.8	62.8	59.2	64.8	60.7	63.1

TABLE III: Cosine similarity between optimal and predicted action (higher is better).

	Mushroom		Barrel		Pot		Mean
	S	L	S	L	S	L	
ID	0.96	1.00	0.98	0.99	0.98	1.00	0.99
OOD	0.62	0.70	0.49	0.72	0.52	0.70	0.63

D. Action Selection

Settings Finally, we evaluate the action selection performance with the cosine distance metric to see if the selected action enables the robot to move towards a more stable configuration, i.e., increases the contact surface area by moving towards the center of mass position of the underlying object.

Results and analyses Table III demonstrates that the policy almost always determines to move toward the near-optimal direction, which can be found from the average cosine similarity being 0.99. In OOD settings, the average similarity score is 0.63, which approximately has an error of 40 degree. In the future, we plan to evaluate these functionalities in the real system to see if these errors are acceptable or not.

VI. Conclusion

Designing systems that can interpret and disentangle useful contact information from observed tactile measurements could be the key to performing precise and fine manipulation. We proposed a framework for estimating extrinsic contact patches from tactile and force-torque measurements for a stacking task with highly irregular shapes. We evaluated the method in a real system and showed reasonable stability estimation performance. In the future, we would like to improve the performance by training on a wider variety of objects and relaxing the assumption of the known geometry so that the trained model can be used for the stacking task with arbitrary objects.

References

[1] M. Riedmiller, R. Hafner, T. Lampe, M. Neunert, J. Degrave, T. van de Wiele, V. Mnih, N. Heess, and J. T. Springenberg, "Learning by playing solving sparse reward tasks from scratch," in *Proceedings of the 5th Conference on Robot Learning (CoRL)*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 4344–4353. [Online]. Available: https://proceedings.mlr.press/v80/riedmiller18a.html

[2] Y. Zhu, Z. Wang, J. Merel, A. Rusu, T. Erez, S. Cabi, S. Tunyasuvunakool, J. Kramár, R. Hadsell, N. de Freitas, and N. Heess, "Reinforcement and imitation learning for diverse visuomotor skills," in *Proceedings of International Conference on Learning Representations (ICLR)*, 2018.

[3] S. Cabi, S. G. Colmenarejo, A. Novikov, K. Konyushkova, S. Reed, R. Jeong, K. Zolna, Y. Aytar, D. Budden, M. Vecerik, O. Sushkov, D. Barker, J. Scholz, M. Denil, N. de Freitas, and Z. Wang, "Scaling data-driven robotics with reward sketching and batch reinforcement learning," 2020.

[4] L. Hermann, M. Argus, A. Eitel, A. Amiranashvili, W. Burgard, and T. Brox, "Adaptive curriculum generation from demonstrations for sim-to-real visuomotor control," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 6498–6505.

[5] R. Jeong, Y. Aytar, D. Khosid, Y. Zhou, J. Kay, T. Lampe, K. Bousmalis, and F. Nori, "Self-supervised sim-to-real adaptation for visual robotic manipulation," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 2718–2724.

[6] M. D. Noseworthy, C. Moses, I. Brand, S. Castro, L. P. Kaelbling, T. Lozano-Pérez, and N. Roy, "Active learning of abstract plan feasibility," *Robotics Science and System (RSS)*.

[7] A. X. Lee, C. M. Devin, Y. Zhou, T. Lampe, K. Bousmalis, J. T. Springenberg, A. Byravan, A. Abdolmaleki, N. Gileadi, D. Khosid, C. Fantacci, J. E. Chen, A. Raju, R. Jeong, M. Neunert, A. Laurens, S. Saliceti, F. Casarini, M. Riedmiller, r. hadsell, and F. Nori, "Beyond pick-and-place: Tackling robotic stacking of diverse shapes," in *Proceedings of the 5th Conference on Robot Learning (CoRL)*, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, 08–11 Nov 2022, pp. 1089–1131. [Online]. Available: https://proceedings.mlr.press/v164/lee22b.html

[8] F. Furrer, M. Wermelinger, H. Yoshida, F. Gramazio, M. Kohler, R. Siegwart, and M. Hutter, "Autonomous robotic stone stacking with online next best object target pose planning," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 2350–2356.

[9] Y. Liu, J. Choi, and N. Napp, "Planning for robotic dry stacking with irregular stones," in *Field and Service Robotics*, G. Ishigami and K. Yoshida, Eds. Singapore: Springer Singapore, 2021, pp. 321–335.

[10] L. Manuelli and R. Tedrake, "Localizing external contact using proprioceptive sensors: The contact particle filter," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 5062–5069.

[11] S. Kim, D. K. Jha, D. Romeres, P. Patre, and A. Rodriguez, "Simultaneous tactile estimation and control of extrinsic contact," *arXiv preprint arXiv:2303.03385*, 2023.

[12] D. Ma, S. Dong, and A. Rodriguez, "Extrinsic contact sensing with relative-motion tracking from distributed tactile measurements," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 262–11 268.

[13] S. Kim and A. Rodriguez, "Active extrinsic contact sensing: Application to general peg-in-hole insertion," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 10 241–10 247.

[14] C. Higuera, S. Dong, B. Boots, and M. Mukadam, "Neural contact fields: Tracking extrinsic contact with tactile sensing," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 576–12 582.