In-Hand Manipulation of Unknown Objects with Tactile Sensing for Insertion

Chaoyi Pan*, Marion Lepert*, Shenli Yuan, Rika Antonova, Jeannette Bohg

***Abstract*— We present a method to manipulate unknown objects in-hand using tactile sensing without relying on a known object model. In many cases, vision-only approaches may not be feasible; for example, due to occlusion in cluttered spaces. We address this limitation by introducing a method to reorient unknown objects using tactile sensing. It incrementally builds a probabilistic estimate of the object shape and pose during task-driven manipulation. Our approach uses Bayesian optimization to balance exploration of the global object shape with efficient task completion. To demonstrate the effectiveness of our method, we apply it to a simulated Tactile-Enabled Roller Grasper, a gripper that rolls objects in hand while collecting tactile data. We evaluate our method on an insertion task with randomly generated objects and find that it reliably reorients objects while significantly reducing the exploration time.**

I. INTRODUCTION

This work studies how robots can reorient objects in-hand with limited prior knowledge about object shape and using only tactile sensing. Existing works have primarily focused on in-hand manipulation using vision [1]–[6], but vision based approaches struggle when the object is heavily occluded. Tactile sensing does not suffer from this problem, but object reorientation with tactile sensing is challenging because tactile data gives information about the object shape for only a small contact patch area. In addition, the object may have only a limited set of features that the tactile sensor can detect to distinguish between different locations on the object, and the object may shift slightly in-between tactile readings, which increases uncertainty regarding the location of these readings.

We seek to overcome some of these challenges by leveraging the Tactile-Enabled Roller Grasper [7]. This gripper rolls objects in hand and continuously collects tactile data on the surface of the rollers. Staying in contact with the object reduces uncertainty between tactile measurements, and enables us to piece together a sequence of local contact patches into a global estimate of the object's shape. Given this gripper, we propose a method to reorient unknown objects by incrementally building a probabilistic estimate of the object's shape during manipulation. Our method leverages Bayesian optimization [8] to strategically trade off exploration of the global object shape with efficient task completion. We demonstrate our approach on a simulated Tactile-Enabled Roller Grasper as shown in Fig. 1.

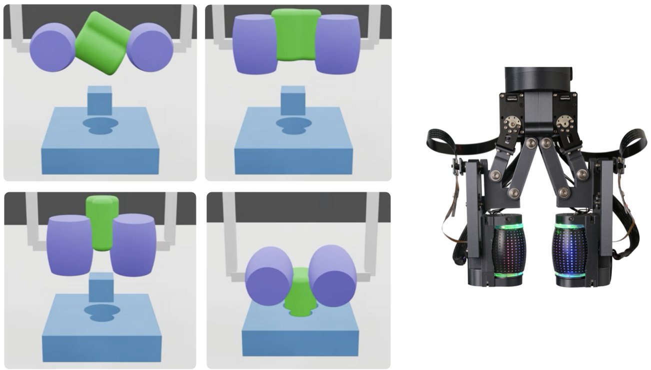

Fig. 1: Left: Simulation of the Tactile-Enabled Roller Grasper demonstrating a sequence of exploration steps that lead to successful insertion. Right: The Tactile-Enabled Roller Grasper that we simulate and that inspired our work.

We evaluate our approach on an insertion task. Insertion tasks are ubiquitous in many environments, such as assembly tasks in factories, dense box packing in warehouses, and plugging cables in the home. As a result, insertion tasks continue to be heavily studied in robotics [9]–[15]. In this paper, we focus on finding the correct object orientation that will allow the object to fit into a target hole. The robot has access to a parameterization of the hole's contour, which gives the robot a well defined reorientation target without revealing the 3D-shape of the object, ensuring that our assumption that the robot is working with an unseen object holds. We evaluate our method in simulation on a set of randomly generated objects and find that our method reliably completes the insertion task while significantly reducing the exploration time needed to do so.

II. APPROACH

Our approach consists of three parts: shape estimation, task oriented exploration, and in-hand reorientation. Shape estimation and task oriented exploration are performed iteratively before the final in-hand reorientation.

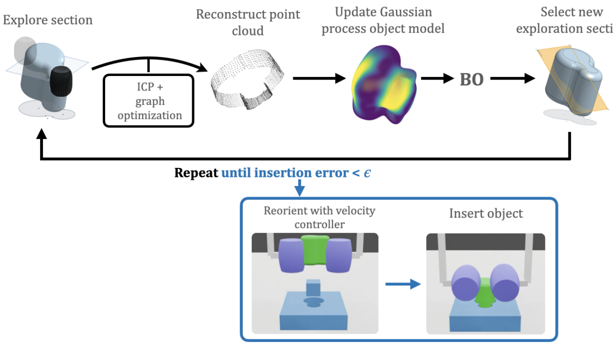
Explore section
ICP + graph optimization
Reconstruct point cloud
Update Gaussian process object model
BO
Select new exploration section
Repeat until insertion error < ϵ
Reorient with velocity controller
Insert object

Fig. 2: A high-level overview of our method.

*Authors contributed equally. Chaoyi Pan is with Tsinghua University. Marion Lepert, Shenli Yuan, Rika Antonova, and Jeannette Bohg are with Stanford University [lepertm, shenliy, rika.antonova, bohg]@stanford.edu. Toyota Research Institute provided funds to support this work. We are grateful to Shaoxiong Wang for help with the Roller Grasper. Rika Antonova is supported by the National Science Foundation grant No.2030859 to the Computing Research Association for the CIFellows Project.

A. *Local shape estimation from tactile images*

To estimate object shape, the Roller Grasper rolls the object in-hand and gathers joint position, $X \in \mathbb{R}^7$, and a tactile height map, $I \in \mathbb{R}^{W\times H}$, at each time step.

We approximate the shape of the object by using the joint position data to estimate the transformation between the depth maps. Because the object occasionally slips on the roller surface, these estimates are noisy. We use ICP between local pairs of depth maps to reduce this noise and improve the fidelity of the reconstruction.

Additionally, once the object has rotated at least 180 degrees, the rollers encounter regions of the object that have already been scanned by the other roller. When we detect this loop closure, we use graph optimization [16] to align the two sets of point clouds.

B. *Global shape representation with Gaussian processes*

Because the proprioception and tactile data collected by the Roller Grasper only give us partial information about the object shape, we use a Gaussian process (GP) to build a probabilistic estimate of the overall shape. We parameterize the surface of the object by spherical coordinates θ, ϕ, r. We use a GP to model the function $f(\theta,\phi) = r$ that represents the distance from the center of the object to its surface along a ray parameterized by θ, ϕ. This GP model lets us predict the mean $\bar{f}(\boldsymbol{x}_*)$ and variance $\mathbb{V}[f(\boldsymbol{x}_*)]$ of the distance to the object's surface along any 'query' ray $\boldsymbol{x}_* := (\theta_*, \phi_*)$.

Formally, $\mathcal{GP}\big(m(\cdot), k(\cdot,\cdot)\big)$ is a model defined by a prior mean function $m(\cdot)$ and kernel function $k(\cdot,\cdot)$. The prior $m(\cdot)$ is usually taken as zero. The kernel encodes similarity between inputs: a large $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ implies that observing the value $r_i = f(\boldsymbol{x}_i)$ for input $\boldsymbol{x}_i$ would have a large influence on our estimate for $f(\boldsymbol{x}_j)$ for input $\boldsymbol{x}_j$. We can compute the posterior mean and variance using:

$$\bar{f}_* := \bar{f}(\boldsymbol{x}_*) := \boldsymbol{k}_*^T (K + \sigma_n^2 I)^{-1} \boldsymbol{r}, \tag{1}$$

$$\mathbb{V}[f_*] := \mathbb{V}[f(\boldsymbol{x}_*)] := k(x_*, x_*) - \boldsymbol{k}_*^T (K + \sigma_n^2 I)^{-1} \boldsymbol{k}_* \tag{2}$$

We obtain $\boldsymbol{r}, K, \boldsymbol{k}_{\boldsymbol{*}}$ from the N points of the object's point cloud $\{\boldsymbol{x}_i := (\theta_i, \phi_i), r_i\}_{i=1..N}$ reconstructed so far. $\boldsymbol{r} \in \mathbb{R}^N$ is a vector with distances from the object's center to its surface (with entries $r_{i,i=1..N}$). $K \in \mathbb{R}^{N \times N}$ is a matrix with entries $k(\boldsymbol{x}_i, \boldsymbol{x}_j)_{i,j=1..N}$; $\boldsymbol{k}_* \in \mathbb{R}^N$ is a vector with entries $k(\boldsymbol{x}_*, \boldsymbol{x}_i)$.

We use the Squared Exponential kernel function and leverage the GP marginal likelihood to estimate the kernel hyperparameters and the noise parameter σ_n automatically. See [17] for further details. Fig. 4 shows an example of incrementally building an object's shape estimate with a GP.

C. *Task Oriented Exploration*

We use *Bayesian Optimization* (BO) to guide our exploration of the object shape in order to efficiently identify regions of the object that are important for completing the insertion task. We construct a task oriented *acquisition* function for BO that selects the next target cross section of the object to explore during in-hand manipulation. The aim is to select targets to reduce the uncertainty over the object shape globally, but focus on the promising regions and avoid over-exploring parts of the object unlikely to be useful for successful insertion. In contrast to related work that focuses on uniformly minimizing uncertainty for overall shape reconstruction [18], [19], we use the task objective to achieve targeted exploration. Below we describe how we construct the task oriented acquisition function for BO.

We represent the orientation of the object with a discrete set of euler angles α, β, γ. Our goal is to evaluate the likelihood of insertion for each orientation. We start with a probabilistic estimate of the object model. This model is a *Gaussian Process* (GP) described in Section II-B, which allows us to obtain estimates along any ray θ, ϕ of the distance to the object surface $R(\theta,\phi) \sim \mathcal{N}(\bar{f}, \mathbb{V}[f])$ in a spherical coordinate system. For each orientation, we decompose the GP object model into horizontal cross sections of uniform width parameterized by height l as shown in Fig. 3.

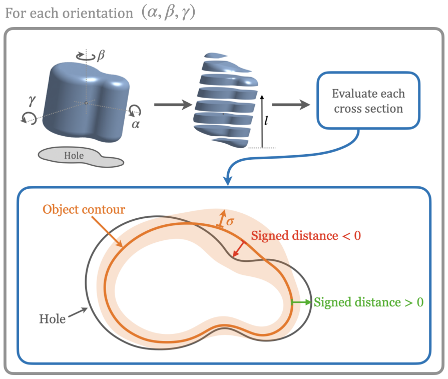

Fig. 3: Overview of our object evaluation: For each orientation (α, β, γ), we split the object into a set of horizontal cross sections of uniform width, parameterized by height l. Each cross section is assigned a probabilistic score using Monte Carlo sampling, reflecting the expected interference between the hole and the cross section. This interference is measured as the minimum signed distance between the hole's contour and the sampled object contour.

Next, we evaluate the likelihood that each cross section will fit into the hole. We start by projecting the GP's cross section to the x-y plane. We parameterize the object's projected contour and the hole's contour with polar coordinates $\theta, r_{\text{obj contour}}(\theta)$ and $\theta, r_{\text{hole}}(\theta)$ respectively. In order for the object to fit into the hole, we need $r_{\text{obj contour}} < r_{\text{hole}}$ $\forall \theta \in [-\pi, \pi]$. We define each section score as

$$S_{\text{section}}(\alpha, \beta, \gamma, l) = \min_{\theta} r_{\text{hole}}(\theta) - r_{\text{obj contour}}(\theta) \tag{3}$$

This score represents the worst location on the object for fit, and a negative score indicates that the object will not fit through the hole. For a given object orientation, the likelihood that the object will fit into the hole is upper-bounded by the worst section, so we set the orientation score to be

$$S_{\text{ori}}(\alpha, \beta, \gamma) = \min_{l} S_{\text{section}}(\alpha, \beta, \gamma, l) \tag{4}$$

We use Monte Carlo sampling from the GP describing the object's surface, $R(\theta,\phi) \sim \mathcal{N}(\bar{f}, \mathbb{V}[f])$, to estimate the object's contour $r_{\text{obj contour}}(\theta) = \max_{\phi} r(\theta, \phi)\cos(\phi)$ and obtain a probabilistic estimate of these scores.

We use these scores to select the next section of the object to explore. Given that each section is parameterized by α, β, γ, l, we decompose this process into first selecting the best orientation α, β, γ and then selecting the height l.

Intuitively, we trade off between selecting an orientation that is most likely to position the object to fit through the hole according to the current model (exploitation), and the orientation with the most uncertainty about its fit (exploration). We can use the mean and standard deviation of S to construct the following acquisition function, based on the *Upper Confidence Bound* (UCB) function [20]:

$$UCB(\alpha, \beta, \gamma) = \mu_{S_{\text{ori}}}(\alpha, \beta, \gamma) + \lambda \sigma_{S_{\text{ori}}}(\alpha, \beta, \gamma), \quad (5)$$

where λ is a hyperparameter that controls the preference between exploitation and exploration. To select the next orientation, we maximize Equation 5 to obtain $\alpha^*, \beta^*, \gamma^*$.

Next, we choose the horizontal section of the object in the selected orientation $(\alpha^*, \beta^*, \gamma^*)$. Here, we trade off between selecting the horizontal section that is most likely to overlap the hole (exploitation) and the horizontal section with the most uncertainty (exploration). This ensures that we tend to examine the 'worst' sections of our 'best' orientations. This is motivated by the need to examine the sections in the 'best' orientation that are most likely to collide with the hole and cause task failure. If such 'worst' sections still fit, the entire object in this orientation is likely to fit into the hole. Formally, we select the parameter l^* (which defines the placement of the section to consider on the vertical axis) by maximizing the following function:

$$\begin{aligned} UCB_{\alpha^*, \beta^*, \gamma^*}(l) = &- \mu_{S_{\text{section}(\alpha^*, \beta^*, \gamma^*)}}(l) \\ &+ \lambda \sigma_{S_{\text{section}(\alpha^*, \beta^*, \gamma^*)}}(l) \end{aligned} \quad (6)$$

D. In-hand reorientation

We use the velocity controller from [3] to determine the rollers' pitch angles $\theta_{L,\text{pitch}}, \theta_{R,\text{pitch}}$ and rollers' angular velocities $\omega_{L,\text{roll}}, \omega_{R,\text{roll}}$ that are necessary to achieve the desired angular velocity of the object. The controller assumes that the object is a sphere, but we find that by adding compliance to the opening of the Roller Grasper, we can reorient a broader set of object shapes.

Due to torsional friction between the rollers and the object, when the roller grasper changes its pitch angle, the object may inadvertently rotate with the roller. This rotation is undesirable because it is hard to control the effects of torsional friction, and it prevents changing roller orientation relative to the object. To mitigate this, we use extrinsic dexterity [21] and lightly press the object against an external surface when changing the pitch angle of the rollers to prevent object rotation.

III. EXPERIMENTAL EVALUATION

A. Simulated task oriented exploration

To evaluate our BO-based task guided exploration strategy, we measure the insertion task success rate and number of required exploration steps using 15 objects and 72 initial grasping points per object. An exploration step consists of a complete rotation of the object along a selected cross section.

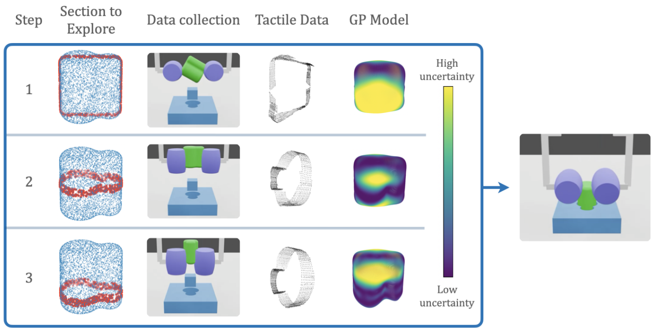

Fig. 4: A demonstration of the reorientation procedure, where each row corresponds to one exploration step. The first column shows the section the roller selects to explore. The second column shows the simulated Roller Grasper rolling the object in hand along the selected section. The third column demonstrates the collected point cloud from the tactile images after using ICP and graph optimization. The fourth column shows the probabilistic object model described by the GP and trained from the collected tactile data.

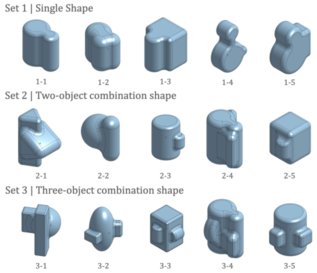
Set 1 | Single Shape
1-1
1-2
1-3
1-4
1-5
Set 2 | Two-object combination shape
2-1
2-2
2-3
2-4
2-5
Set 3 | Three-object combination shape
3-1
3-2
3-3
3-4
3-5

Fig. 5: The object set used for evaluation.

Simulation environment. The simulation environment is built in PyBullet [22]. First, an object is instantiated in-hand with a random orientation sampled from a uniform distribution. Next the rollers roll the object in-hand and collect tactile data. When a closed loop is detected, the roller stops and proceeds to the next section determined by the selected algorithm. The exploration process stops when the insertion error is less than 3mm.

Object set. Fig. 5 shows the 15 different objects we used for evaluation, generated by randomly combining either one, two, or three randomly-transformed basic 3D shapes. The target hole shape is generated by projecting the object onto a plane and verifying that the object can fit in the hole in only one orientation. This ensures that our algorithm must be accurately reconstructing the object shape to solve the task. The maximum object width is 5cm.

Methods we compare.

1) Our method that uses Bayesian optimization to trade off between exploration and task completion, with λ optimized over a hold-out set of objects ($\lambda = 500$).
2) Random: selects sections for exploration from a uniform random distribution
3) Exploit-only: without considering uncertainty, selects the object orientation that is most likely to allow

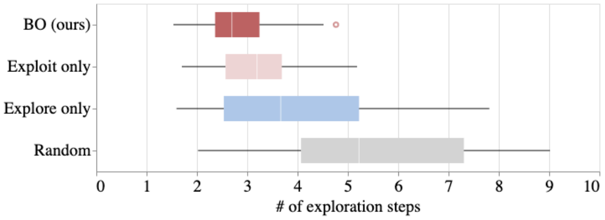

Fig. 6: Number of exploration steps required for successful object insertion using different algorithms in simulation. The boxplot indicates the 25th percentile, median, and 75th percentile over the 15 objects tested.

insertion and picks the section in that orientation that is most likely to cause a failure. This is equivalent to our BO algorithm with $\lambda = 0$.

4) Explore-only: selects the object orientation and section with the most uncertainty. This is equivalent to our BO algorithm with $\lambda \to \infty$.

The maximum exploration time allowed is 10 steps.

Quantitative results. Each result is evaluated over 15 objects with 72 initial grasping points. All methods are able to perform with a high success rate given enough exploration steps $> 98\%$. However, our BO-based exploration strategy consistently requires fewer exploration steps (Fig. 6). Over the randomly-generated set of objects, our BO approach performs better than exploit-only ($p < 0.01$), explore-only ($p < 0.01$), and the random baseline ($p < 0.001$). All p-values were computed using the paired-samples t-test.

Shape reconstruction. While shape reconstruction is *not* the explicit objective of our algorithm, we still demonstrate that qualitatively, we end up recovering a meaningful approximation of global object shape as an auxiliary benefit of minimizing the task-driven objective (Fig. 8).

B. Sensitivity analysis for noise

Noise in the reconstructed point cloud. Noisy depth images and a noisy prior estimate on the object's motion due to slipping could cause poor object reconstruction results. To evaluate the tolerance to poor object reconstruction, we add i.i.d. zero-mean Gaussian noise to the reconstructed point cloud before fitting the GP model. As shown in Fig. 7 (left), while the success rate deteriorates in all methods as noise increases, our method outperforms others in the low-to-medium noise regimes. For high-noise regime our method performs similarly to the exploit-only baseline.

Noise in the depth image from the tactile sensor To evaluate the tolerance to noisy depth images, we add i.i.d. zero-mean Gaussian noise to the simulated depth images used for ICP reconstruction. Unlike adding noise to the reconstructed model, adding noise to depth images makes it harder for ICP to reconstruct the object model. Fig. 7 (right) demonstrates that while adding noise to the depth images affects the performance of all methods, our method has a significant advantage in the low-to-medium noise regimes. When the standard deviation of the noise is smaller than 4 mm, our method can still reach a success rate above 80%. Higher noise levels can lead to a noticeable performance drop, at which point our method achieves comparable performance to our baselines. This is because our method depends on making informed guesses of where to explore next based on the model's current estimate of the object shape. With high levels of noise, the shape estimate becomes very inaccurate, and our algorithm no longer has enough information to make an informed guess of where to explore next.

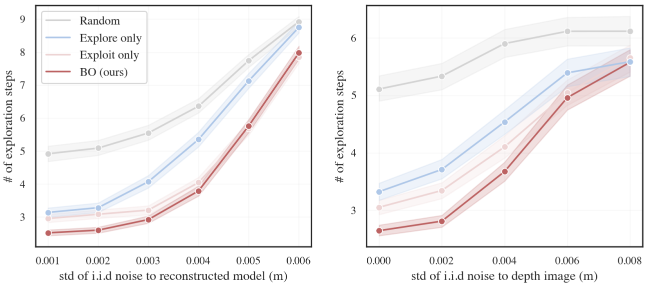

Fig. 7: Steps to finish the exploration with i.i.d. noise. Each bar is evaluated over 5 objects with 72 different initial grasping points.

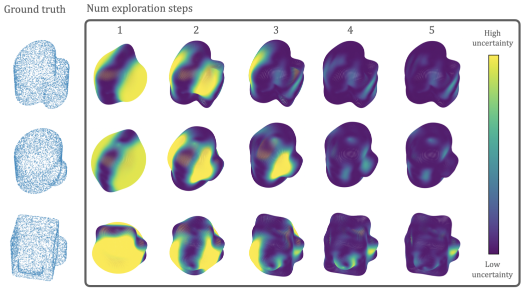

Fig. 8: Qualitative visualization of the shape reconstruction.

Hardware implications Our simulation study provides instructive feedback for deployment of our approach on the Tactile-Enabled Roller Grasper hardware. The Tactile-Enabled Roller Grasper used in this study is a novel gripper that is still under development. We empirically estimate the approximate noise of the depth image to be at least 7mm. Fig. 7 demonstrates that at this level of precision, the BO algorithm's benefit over a simpler strategy is limited. Our primary takeaway is that the gripper's sensor noise needs to be reduced by replacing the 3D printed plastic links with stiffer materials and reducing the backlash in the joints to improve the gripper's proprioception. Additionally, adding a third roller is needed to enable the rollers to change pitch angle without experiencing high shear forces that degrade the GelSight surface. The existing two roller design requires a high normal force when changing pitch angle to avoid dropping the object which causes high shear forces.

IV. CONCLUSION

We present a method to reorient unknown objects with tactile sensing that does not rely on vision. We perform in-hand simultaneous 3D shape reconstruction and localization, and outline an efficient strategy based on Bayesian optimization to select regions of the object to explore to ensure quick task completion. We demonstrate the efficacy of this method on the insertion task, suggesting its possible broad utility in tactile manipulation.

References

[1] OpenAI: Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, et al. Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39(1):3–20, 2020.

[2] Nikhil Chavan-Dafle, Rachel Holladay, and Alberto Rodriguez. In-hand manipulation via motion cones. *arXiv preprint arXiv:1810.00219*, 2018.

[3] Shenli Yuan, Lin Shao, Connor L Yako, Alex Gruebele, and J Kenneth Salisbury. Design and control of roller grasper v2 for in-hand manipulation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9151–9158. IEEE, 2020.

[4] Silvia Cruciani, Balakumar Sundaralingam, Kaiyu Hang, Vikash Kumar, Tucker Hermans, and Danica Kragic. Benchmarking in-hand manipulation. *IEEE Robotics and Automation Letters*, 5(2):588–595, 2020.

[5] Andrew S. Morgan, Kaiyu Hang, Bowen Wen, Kostas Bekris, and Aaron M. Dollar. Complex in-hand manipulation via compliance-enabled finger gaiting and multi-modal planning. *IEEE Robotics and Automation Letters*, 7(2):4821–4828, 2022.

[6] Tao Chen, Jie Xu, and Pulkit Agrawal. A system for general in-hand object re-orientation. In *Conference on Robot Learning*, pages 297–307. PMLR, 2022.

[7] Shenli Yuan. *Robot in-hand manipulation using Roller Graspers*. PhD thesis, Stanford University, 2022.

[8] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando de Freitas. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.

[9] Herman Bruyninckx, Stefan Dutre, and Joris De Schutter. Peg-on-hole: a model based solution to peg and hole alignment. In *Proceedings of 1995 IEEE International Conference on Robotics and Automation*, volume 2, pages 1919–1924. IEEE, 1995.

[10] Siddharth R Chhatpar and Michael S Branicky. Search strategies for peg-in-hole assemblies with position uncertainty. In *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, volume 3, pages 1465–1470. IEEE, 2001.

[11] Hyeonjun Park, Jaeheung Park, Dong-Hyuk Lee, Jae-Han Park, Moon-Hong Baeg, and Ji-Hun Bae. Compliance-based robotic peg-in-hole assembly strategy without force feedback. *IEEE Transactions on Industrial Electronics*, 64(8):6299–6309, 2017.

[12] Karl Van Wyk, Mark Culleton, Joe Falco, and Kevin Kelly. Comparative peg-in-hole testing of a force-based manipulation controlled robotic hand. *IEEE Transactions on Robotics*, 34(2):542–549, 2018.

[13] Hee-Chan Song, Young-Loul Kim, and Jae-Bok Song. Automated guidance of peg-in-hole assembly tasks for complex-shaped parts. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4517–4522, 2014.

[14] Te Tang, Hsien-Chung Lin, Yu Zhao, Yongxiang Fan, Wenjie Chen, and Masayoshi Tomizuka. Teach industrial robots peg-hole-insertion by human demonstration. In *2016 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 488–494, 2016.

[15] Michelle A. Lee, Yuke Zhu, Krishnan Srinivasan, Parth Shah, Silvio Savarese, Li Fei-Fei, Animesh Garg, and Jeannette Bohg. Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8943–8950, 2019.

[16] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5556–5565, 2015.

[17] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[18] Marten Björkman, Yasemin Bekiroglu, Virgile Högman, and Danica Kragic. Enhancing visual perception of shape through tactile glances. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3180–3186. IEEE, 2013.

[19] Danny Driess, Peter Englert, and Marc Toussaint. Active learning with query paths for tactile object shape exploration. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 65–72. IEEE, 2017.

[20] Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. pages 1015–1022, 07 2010.

[21] Nikhil Chavan Dafle, Alberto Rodriguez, Robert Paolini, Bowei Tang, Siddhartha S Srinivasa, Michael Erdmann, Matthew T Mason, Ivan Lundberg, Harald Staab, and Thomas Fuhlbrigge. Extrinsic dexterity: In-hand manipulation with external forces. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1578–1585. IEEE, 2014.

[22] Erwin Coumans and Yunfei Bai. Pybullet, a python module for physics simulation for games, robotics and machine learning. `http://pybullet.org`, 2016–2021.