UNIVERSITY OF CALIFORNIA SAN DIEGO

A Grasp Analyzer for Stable Manipulation of Object in Tabletop Environments

A dissertation submitted in partial satisfaction of the
requirements for the degree Master of Science in Computer Science

in

Master of Science

by

Zhao Tang

Committee in charge:

Professor Henrik I. Christensen, Chair
Professor Hao Su
Professor Xiaolong Wang

2023

The Dissertation of Zhao Tang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Jiaming Hu for his contribution to my thesis, my knowledge and my skills. This work would not have been possible without his passion and dedication to robotics.

I would also like to acknowledge Professor Henrik I. Christensen for introducing me to this wonderful field as well as advising us throughout this work.

Chapter 4, in part, is being submitted for publish of the material as it may appear in International Conference on Intelligent Robots and Systems(IROS). The dissertation author wrote those paragraphs, and is a major contributor and author to the work described in those paragraphs.

ABSTRACT OF THE DISSERTATION

A Grasp Analyzer for Stable Manipulation of Object in Tabletop Environments

by

Zhao Tang

Master of Science in Computer Science in Master of Science

University of California San Diego, 2023

Professor Henrik I. Christensen, Chair

Human-like manipulation for robots has been an ongoing research effort for decades. Among such efforts, stable pick-and-place is one of the simplest yet impactful tasks. The ability to perform stable pick-and-place tasks, especially with unseen object in cluttered environments, would enable robots to handle common household object and eventually perform more complex tasks in kitchens, labs and offices. This thesis presents a novel grasp analyzer to identify stable sliding and lifting grasp poses for parallel grippers. This grasp analyzer combines learning-based models with physics-based models to ensure performance and robustness. Combined with a multi-modal planner, the grasp analyzer was proved with experiments to work both in simulation and real world.

Chapter 1

Introduction

1.1 Motivation

Robotic manipulation is a fundamental field in robotics. Over the last two decades, robotics arms have become core tools in many industries such as car manufacturing, minimally invasive surgical operations and precision assembly of electronics. However, open-world manipulation tasks still remain a largely unsolved problem[14]. An example of such a task is shown in 1.1. One of the main reasons current systems' performance are unsatisfactory at such tasks is that we expect robots to achieve human level performance in manipulation tasks while current robots are ill-equipped to succeed. Robotic manipulators don't have the suitable mechanical and sensing hardware to succeed in manipulation tasks [17] [19]. In pick-and-place tasks, this issue is apparent. For instance, thanks to our dexterous hands, humans can easily pick up unseen objects in complex environments in a stable manner. We are able to dance around obstacles with a mug of coffee without spilling it. However, most robots use parallel gripper instead of hand-like grippers for their low cost and ease of control. As a result, such grippers provide much less points of contact compared to the human hand, thus limiting the amount of stable grasp poses. This makes stable manipulation challenging. This deficit in hardware calls for creative algorithms tailored for current robotic hardware that both ensures performance and robustness in open-world environments.

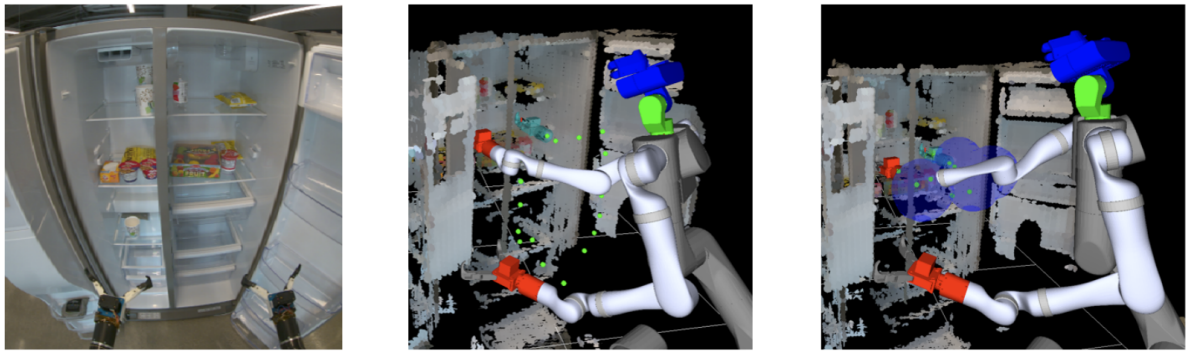


Figure 1.1. An example of a robotic manipulation task in a home environment. The green dots signify the planned path for the end effecter to take. [2]

1.2 Contributions

The main contribution of this thesis is a learning and physically based grasp analyzer that can differentiate between stable lifting and stable sliding poses. The grasp analyzer consists of three parts: a learning-based 6-DoF grasp proposer (pre-trained model, not novel), a learning-based center of mass estimator, and a physics based grasp classifier that determines the quality of each proposed grasp. This grasp analyzer is intended to use with a multi-modal planner [7]. Together, the system enables stable pick-and-place in table-top environments even in cluttered environments containing unseen objects with complicated shapes.

Chapter 2

Background

This section provides background for each of the individual components of the grasp analyzer.

2.1 Center of Mass Estimation

Humans have an inert ability to estimate the center of mass of common objects through vision alone. This ability enables us to manipulate objects stably as we will minimize the torque generated by a grasp. Conversely, center of mass estimation has also been an important part in the classical (non-learning) approach to grasping [8], [20]. [20] proposed a method that uses the gravity equi-effect plane of an object under external force to determine the center of mass of a polyhedron. However, this method requires the robot to tip the object repeatedly, which may not be feasible in cluttered environments. Moreover, since a lot of common objects are not polyhedrons, the application and accuracy of the method may be limited. [8] formulated the process of grasping and finding the center of mass of the object as a reinforcement learning problem. An initial estimate of the CoM (center of mass) is found by finding the centroid of a segmented object's point cloud. Using this CoM, the grasp with the lowest expected torque is executed by the robot and the actual grasp torque is measured by force sensors. This data is then used to re-evaluate the CoM, repeating the procedure until a satisfactory low-torque grasp has been found. This method could lead to irrecoverable failures as a grasp must be executed

to calibrate the center of mass found. More importantly, the centroid of a point cloud is not a reliable estimator for center of mass as the full point cloud of an object is usually not attainable in open-world environments. The centroid of partial point cloud may be misleading, causing slippage during grasp.

2.2 Grasp Pose Proposer

Due to the limited contact points provided by a parallel gripper, finding a suitable grasp pose becomes a challenging problem. In recent years, machine learning has become popular approach for grasp proposers [15], [9], [13]. These learning-based methods are able to generate a set of 6-DoF grasp poses for parallel grippers from raw point cloud data. [15] proposed a CNN that is trained using artificial data. The data consists of sampled grasp poses that are labeled as good and bad grasp poses. [9] proposed a two-part framework where a VAE is trained to sample poses with a point cloud input and an evaluation module is trained to detect and refine the proposed grasps. Contact-GraspNet[13] simplified the representation of grasp poses from 6-DoF to 4-DoF to improve model training. It also used the ACRONYM [4] dataset, a large simulation-based dataset with a vast amount of different object types. Contact-GraspNet was able to achieve state-of-the-art performance on many benchmarks while being lightweight. It consisted of only one network with a Pointnet++ [10] encoder backbone and returned directly usable grasp poses. However, since ACRONYM's grasp poses are classified under a zero-gravity environment, the grasp poses generated by Contact-GraspNet do not consider instabilities such as rotation and slippage when the objects' pose change relative to gravity. This leads to the model proposing grasp poses that can be maintained but are unstable. This is shown in 2.1.

2.3 Grasp Stability Analysis

The notion of stability has not agreed-upon definition [5]. In certain contexts, stability may refer to the existence of an equilibrium in which the contact point of the object and the

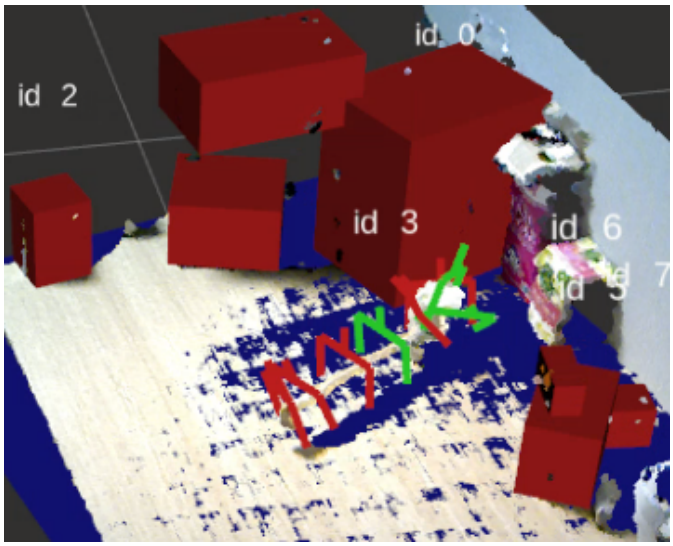


Figure 2.1. Some grasp poses generated by Contact-GraspNet. The red grasp poses are not stable if the object is being lifted, while the green ones are stable lifting grasp poses. The poses are classified by the grasp analyzer proposed.

gripper obey certain friction laws. With hand-like grippers that provide a high number of contact points, there exists a state of stability where the normal forces of contact completely limit the movement of the object. This is referred to as "form closure" [3] and is a very strong stability condition. An object under form closure is completely kinematically constrained by the contacts of the gripper. Thus unless the motors on the gripper fail, no force in any direction would destabilize the object. However, such analysis would require prior knowledge of the object's shape, which is often unobtainable in open-world settings.

Moreover, in the context of parallel grippers with two pins, it is nearly impossible to achieve form closure. Therefore, it is important for parallel grippers to consider the destabilizing forces and how they affect the grasp under different poses. In an isolated environment where the robot is the only actor,gravity is the only destabilizing force. Thus, given the center of mass of the object and the mass of the object, one can perform simple force and torque analysis to determine the stability of a grasp.

Chapter 3

Methodology

The grasp analyzer consists of a grasp proposer, a center of gravity (CoM) estimator, and a grasp classifier 3.1. The first two components are learning-based; the grasp classifier is a physics-based torque analyzer. For the grasp proposer, a pre-trained version of Contact-GraspNet [13] is used. Contact-GraspNet is chosen for its simplicity as well as its state-of-the-art performance in tabletop scenarios. The CoM estimator and the grasp classifier will be discussed in more details in the following sections.

3.1 Center of Mass Estimator

3.1.1 Architecture

The architecture of the center of mass estimator is shown in 3.2. Similar to Contact-GraspNet, the first part of the architecture utilizes a Pointnet++ feature extractor to create per-point features from segmented point cloud. This feature is then passed into three fully connected layers. Note that the fully connected layers do not perform reduction to directly generate a CoM estimate. Instead, they produce a "dense" (per point) estimate of the CoM as proposed by [16]. This has been shown in the pose estimation community to drastically improve the stability and accuracy of estimation from point clouds. In [16], mean shift with consensus is used to aggregate the per-point estimations. However, in the setting of center of mass estimation, using mean shift or RANSAC did not show any improvement in accuracy. Therefore, a simple

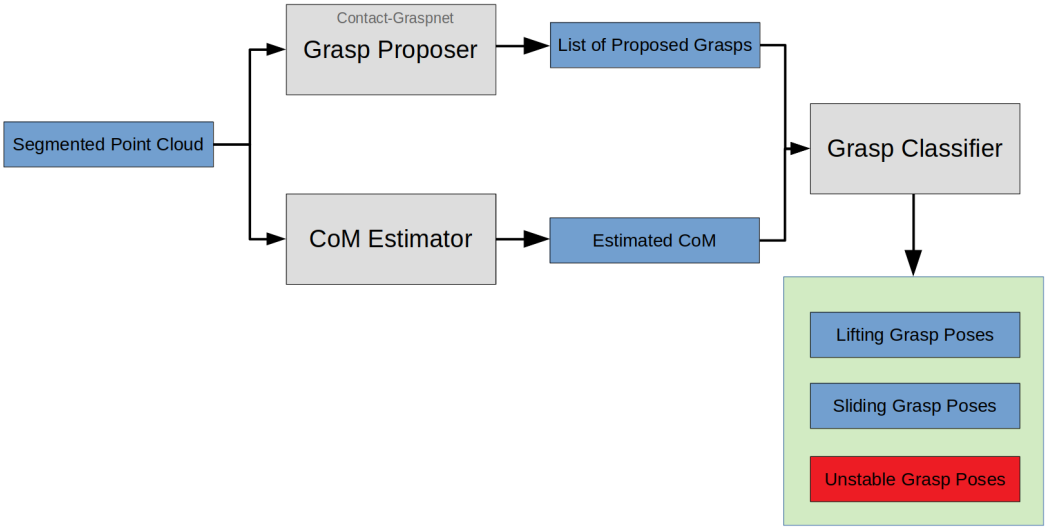


Figure 3.1. The overall pipeline of the grasp analyzer is shown here. The grasp proposer and the CoM estimator take segmented point cloud are input and output a list of proposed grasps and an estimated CoM. The two are then analyzed the grasp classifier to determine which grasps are stable lifting grasps, stable sliding grasps or neither. Note that a grasp can be both a stable lifting grasp and a stable sliding grasp.

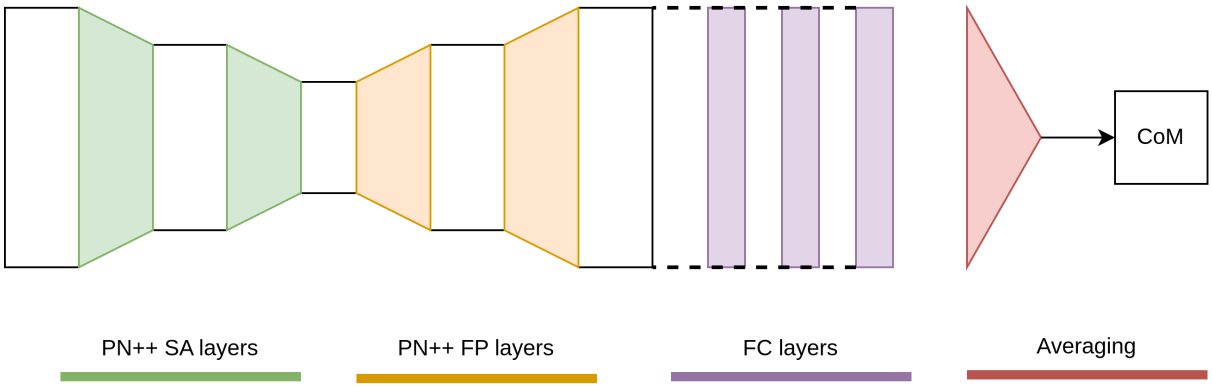


Figure 3.2. The architecture of the center of mass estimator is shown here. The first part of the network is a point cloud feature extractor with three Pointnet++ Set Abstraction layers and three Pointnet++ Feature Propagation layers. The per-point features are then fed into three fully connected layers. As the layers do not reduce in dimensions, they can also be viewed as convolution layers. The fully connected layers produce estimates of the object's center of mass for each point in the point cloud.

averaging is used to aggregate the dense estimates to form the final center of mass estimation.

3.1.2 Point Cloud Augmentation

The aforementioned architecture achieves a satisfactory performance in terms of overall accuracy. However, there are some egregious failure cases such as predicting CoMs that are below the table as shown in 3.3. To remedy these errors, an "imagined" table point cloud is added similar to [11]. In [11], an imaginary hand point cloud is added to improve the model's cognition of the spatial relationship between the hand and the manipulation target. In the case of CoM estimation, the imaginary table surface point cloud serves a similar purpose. It allows the model to get a better sense of the size of the object even under occlusion. It also allows the model to better assess the the placement of the object on the table. Given that the object must be in a stable placement, additional assumptions about the object's distribution of mass can be made to help with center of mass prediction. With the imaginary table surface point cloud added to augment the segmented partial point cloud of the objects, egregious errors with CoM estimations decreases significantly.

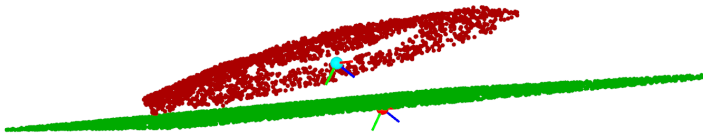


Figure 3.3. An example of estimation failure. The ground truth center of mass if shown in blue while the predicted center of mass is shown in red. The object's observed point cloud is shown in red and the table is shown in green. Because only the partial object point cloud is passed to the model, the model is unsure of the exact dimensions of the object and makes a prediction that overestimates the size of the object.

3.1.3 Data Set and Training

The model is trained using a data set derived from the ACRONYM database [4]. ACRONYM contains center of mass information of ShapeNet [1] objects. ACRONYM's toolkit also contains methods that allows the user to generate tabletop scenes. A total of 1000 scenes are generated by placing a random object in a random stable placement on a table. For each scene, five different camera angles that resemble a robot's view of the tabletop are created randomly. Each camera angle would generate a distinct partial point cloud of the object. Each partial point cloud is augmented with a labeled imagined tabletop point cloud; center of mass and a camera matrix annotations are also included. An example of the five point clouds from a different scenes are shown in 3.4. Of the 5000 samples, 4000 are used to train the model and 1000 are reserved for evaluation.

The model is trained with a learning rate of 0.001 using the ADAM optimizer. The loss function is the average L2 loss of the dense predicted CoM and the ground truth CoM. The final model achieves an average CoM estimation error of 2cm on the evaluation data set. Some example predictions are shown in 3.5.

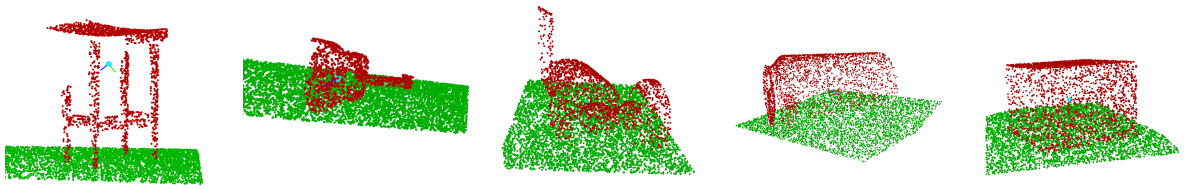


Figure 3.4. Examples of generated data with partial object point cloud in red, augmented table point cloud in green, ground truth center of mass as a blue dot. The axis shown on the ground truth center of mass is aligned with the camera reference frame.

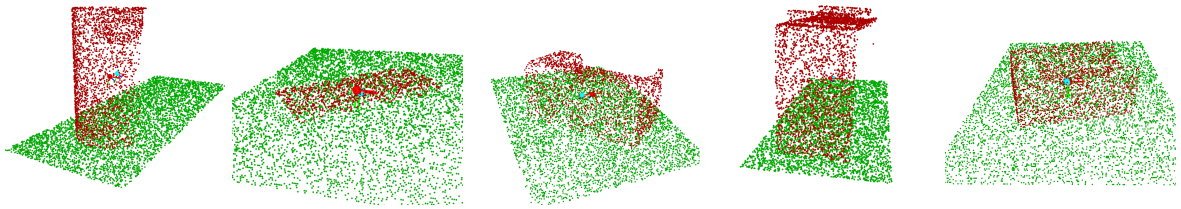


Figure 3.5. Examples of prediction results. The ground true center of mass is shown as a blue dot while the predicted center of mass is shown as a red dot.

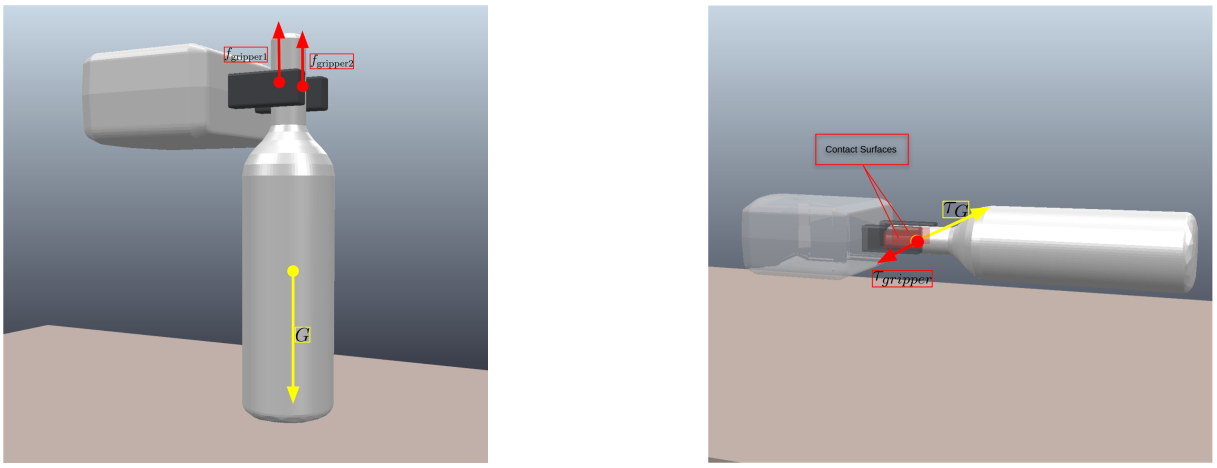
3.2 Grasp Classifier

3.2.1 Assumptions

This work assumes that the parallel gripper has only two pins with powerful motors. This work also assumes that the manipulation task is in isolation, meaning that the robot is the only actor in the task. In such context, the forces acting on the object are gravity and the contact forces exerted by the two pins as well as the resulting friction.

A stable grasp for lifting is defined as when the object is lifted by the grasp, the relative pose between the gripper and the object does not change; vice versa, a stable grasp for sliding is defined as when the object is slid by the grasp, the relative pose between the gripper and the object does not change.

Under this setting, it is possible to classify grasp destabilization as two types: slippage destabilization and rotational destabilization. Slippage destabilization refers to the scenario when the contact friction between the gripper and the object is not sufficient to support the object's



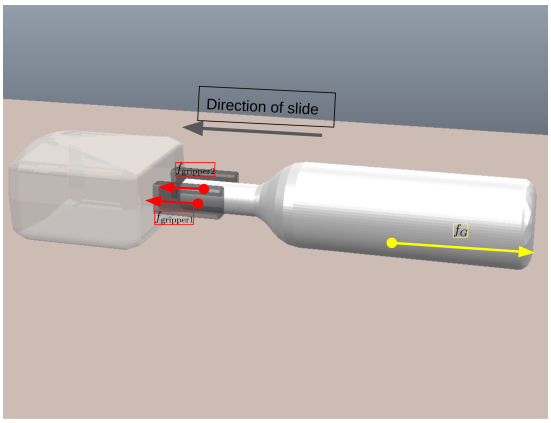
(a) Slippage **(b)** Rotational

Figure 3.6. (a) shows the forces in play for a lifting grasp concerning slippage destabilization. f_{gripper1} and f_{gripper2} are the static friction exerted by each of the two gripper pins; G is the gravitational force experience by the object. (b) shows the torques in play for a lifting grasp concerning rotational slippage. By choosing the center of grasp as pivot, $\tau_{gripper}$ is the static torque created by the contact surface between the gripper and the object; τ_G is the gravitational torque experienced by the object.

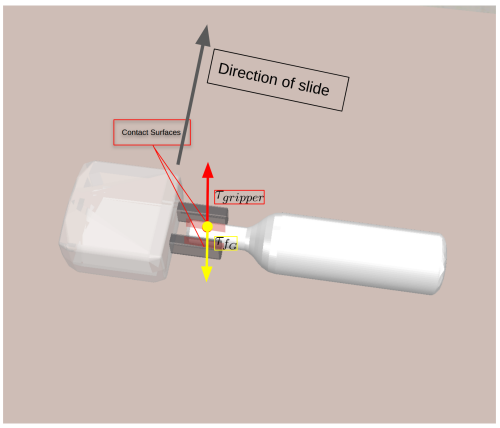
weight. Rotational destabilization refers to the the scenario when the static torque between the gripper and the object is not sufficient to balance out the object's gravitational torque. The two types of destabilization are illustrated for the lifting case in 3.6 and for sliding case in 3.7.

It is assumed that slippage destabilization does not happen for any of the grasps proposed. This assumption is based on three reasons. First, grasps generated by Contact-GraspNet are usually immune to slippage as the network is trained to prevent such destabilization. Second, the motors on the gripper are usually very powerful. Thus, in practice, the maximum friction and normal forces generated by the gripper usually far exceed the gravitational forces and the friction between the object and the table. Finally, if slippage destabilization is possible, it usually means that the object is too heavy or has a very high friction coefficient for the gripper to handle. If that is the case, grasp analysis become a moot point as the robot simply lacks enough power to manipulate the object.

Given the above assumptions, the focus of the grasp classifier is therefore to analyze the torques in play to prevent rotational destabilization during manipulation.



(a) Slippage



(b) Rotational

Figure 3.7. (a) shows the forces in play for a sliding grasp concerning slippage destabilization. The object is being slid to the left. f_{gripper1} and f_{gripper2} are the static friction exerted by each of the two gripper pins; f_G is the friction caused by gravitational force experience by the object. (b) shows the torques in play for a sliding grasp concerning rotational slippage. The object is being slid upwards. By choosing the center of grasp as pivot, $\tau_{gripper}$ is the static torque created by the contact surface between the gripper and the object; τ_{f_G} is the torque experienced by the object due to the friction between it and the table.

3.2.2 Model and Notions

An accurate model for torque analysis would be intractable. For instance, pivot point selection in non-equilibrium systems become complicated. Therefore, most approaches such as [8] simply consider grasps with smaller distances to the center of mass as stable grasps. However, such models neglect two-pinned parallel grippers generate vastly different static torques in different axis. For instance, the maximum static torque generated by the gripper in 3.6b would be much smaller than the in 3.7b, and this would lead to a rotation in 3.6b while the grasp in 3.7b would remain stable.

To model this property while keeping calculations tractable, maximum static torque is defined as three separate values on each of the three axis for a Fetch [18] gripper 3.8. The pivot point is selected as the center of grasp, denoted as the origin in 3.8. As slippage destabilization is unlikely to happen, this pivot point is a good approximation to the actual pivot point of the object when motion is relatively low-velocity. The maximum static torque on each axis is denoted

as τ^x_{max}, τ^y_{max} and τ^z_{max} respectively. The torque generated by either the gravitational force or friction between object and table is denoted as τ_{rotation}. This torque can be found with a simple cross product:

$$\begin{aligned}\tau_{\text{rotation}} &= L_{\text{rotation}} \times F_{\text{rotation}} \\ &= (P_{\text{rotation}} - P_{\text{GraspCenter}}) \times F_{\text{rotation}}\end{aligned}$$

Where the L_{rotation} is the lever arm and F_{rotation} is the rotational force (gravitation force or friction). L_{rotation} is found by the displacement between the object's grasp center and the rotational force's point of exertion.

The rotational torque is then decomposed into the gripper's reference frame via a vector-matrix multiplication

$$\begin{bmatrix} \tau^x_{\text{rotation}} \\ \tau^y_{\text{rotation}} \\ \tau^z_{\text{rotation}} \end{bmatrix} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{bmatrix} \cdot \tau_{\text{rotation}}$$

where $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{bmatrix}$ represents the orthogonal normal matrix of the gripper's reference frame. The rotation torque in each axis is then compared to the maximum static torque in each axis to determine the grasp pose's stability. If none of the rotational torques in any direction exceeds the maximum static torque, the pose is deemed stable. Otherwise, it is deemed unstable. An example is shown in 3.9.

3.2.3 Additional Details

For lifting grasp poses, since gravitational force only has one direction, torque analysis only needs to be performed once. For sliding grasp poses, since the frictional force's direction depends on the direction the object is being slid, a single torque analysis is not sufficient. For each sliding pose, a number of directions are sampled uniformly from 2π and each direction

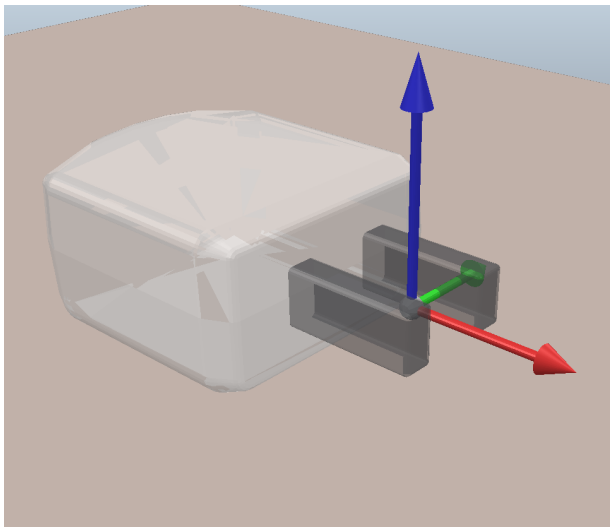
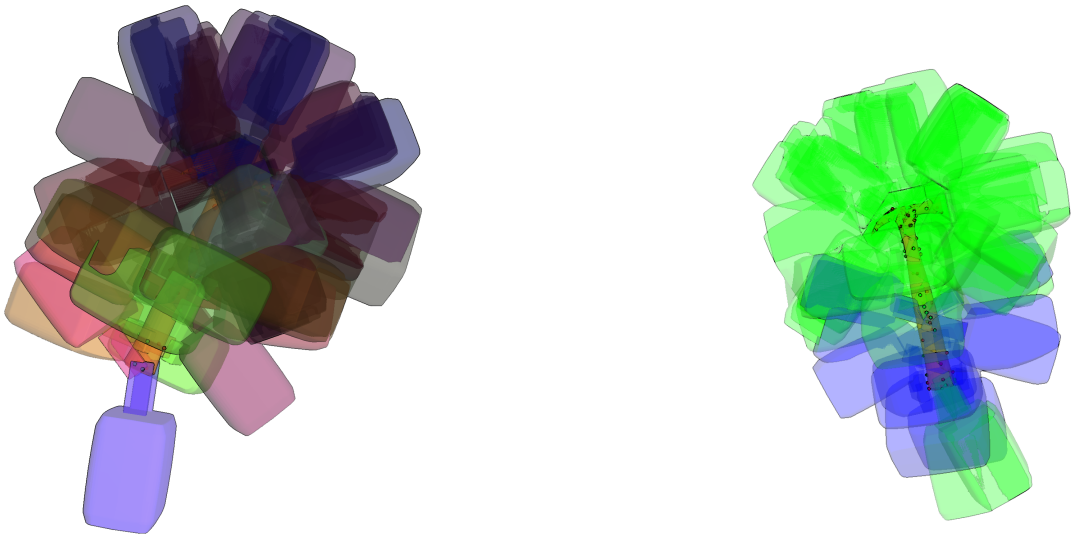


Figure 3.8. The reference frame of a Fetch gripper is shown here. The RGB axis illustrates the maximum static torque on the XYZ axis. A fetch gripper would offer the most static torque on the Z axis and the least static torque on the Y axis.

is analyzed. A sliding pose is stable only if all directions have stable torques. The sampling density is a parameter that can be tuned. For sliding poses analysis, the exact point on which of the friction is exerted is also unknown. Ideally, it can be found by integrating over the mass distribution and the friction coefficient of the object. In practice, this point can be estimated as the projection of the center of mass of the object onto the table. This is true if the object's friction coefficient is even across the contact surface.

In practice, the mass of the object, the friction coefficient and the maximum static torques of the gripper are unknown. Therefore, it is impossible to give exact classification. Instead, in practice, the grasp poses are first clustered to reduce their number. Then, each pose is ranked based on their torque analysis results. The top ranked poses are classified as stable poses. The clustering process allows for a diverse set of grasps to be recommended. Both clustering and ranking parameters can be tuned.



(a) Torques **(b)** Classifications

Figure 3.9. This example shows the torque analysis and subsequent classification of grasp poses for a hammer object. Each grasp pose is shown as a transparent gripper. (a) shows the torques in gripper XYZ frame. The three torques are visualized as RGB values in each grasp pose. (b) shows the subsequent classification made based on the torques in each axis. The green grasp poses are stable lifting poses while blue ones are not. Notice that two grasp poses at the end of the handle are still classified as stable even though they generate a large torque.

Chapter 4

Experiments and Results

No individual experiments were performed to assess the performance of the stand-alone grasp analyzer. Rather, the grasp analyzer was integrated with a multi-modal planner [7] to perform pick-and-place tasks both in-simulation and in-real-world experiments. Since both the grasp analyzer and the multi-modal planner has to perform to succeed in a task, the experiments would reflect the performance of the grasp analyzer.

The robot is asked to pick up an object from a cluttered table in a stable manner. To ensure stability during pickup and placement, the multi-modal planner uses stable sliding grasp poses to relocate the object to a position where a stable lifting pose is feasible; then, the planner uses the stable lifting pose to pick up and place the object. For each scene and object, a set of trials are performed using the grasp analyzer enabled multi-modal planner and another set of trials are performed using only Contact-GraspNet and an RRT-based planner. The stability of each solution is then compared.

4.1 Simulation experiments

4.1.1 Setup

The simulation experiments are performed in Coppeliasim [12] on a simulated Fetch platform. A total of 12 different objects in 12 distinct scenes are performed with the Bullet physics engine. The scene consists of multiple object place on a table, creating a cluttered

environment. A Fetch platform, having a parallel gripper with 100 N of grasping force, is positioned at some distance in front of the table. In the first 6 scenes, the target objects are deliberately place in locations where stable lifting poses are initially infeasible. In the remaining 6 scenes, objects are placed randomly.

4.1.2 Experiment Trials and Stability Metric

A trial in the simulation experiment involves the robot grasping the object and then removing the supporting table from the simulation. After the table has been removed fro 5 seconds, a stability core is measured. *Grasping Stability* is measured by the inverse of the object's rotational change relative to the gripper after the table has been removed: stability $= \frac{1}{\Delta R}$. A stable grasp would result in a low object pose change relative to the gripper after grasp, and thus a higher stability score. We also measure the final height of the object after the table has been removed. If the height of the object is less than the height of the removed table, then the object is considered to be dropped. In this case, the stability score is set to 0, indicating a failure. For each object, 10 total trials were run: 5 trials on the Contact-GraspNet and RRT based baseline system and 5 trials on the grasp analyzer and multi-modal task planner based system. The performances of the systems are judged based on the average stability score in a scene.

4.1.3 Results

Results are shown in Table 4.1. In all of the scenes, our method achieved on-par or higher stability scores compared to the baseline method. For certain scenes and objects, the baseline was unable to complete the grasping task entirely, resulting in a stability score of 0. In contrast, our grasp analyzer and multi-modal planner were able to identify and execute stable grasps in all of the scenes. This proves that the grasp analyzer's judgments of stable sliding and lifting grasps are sound.

Table 4.1. Simulation Experiment Result

object name	object mass(kg)	avg. stability (higher is better)	
		baseline	our method
hammer	0.10	1.04	**9.06**
pan	0.16	1.71	**5.91**
candy bar	0.16	0.00	**32.6**
microphone	0.10	23.5	**41.6**
level	0.51	0.00	**6.23**
wrench	0.12	0.00	**13.5**
tissue box	0.80	3.92	**4.36**
cereal	0.12	5.94	**25.3**
can	0.10	19.3	**68.6**
caliper	0.18	5.06	**8.29**
dispenser	0.25	0.71	**35.9**
remote	0.23	0.35	**5.11**

4.2 Real World Experiments

4.2.1 Setup

We conducted real-world experiments on three objects - a hammer, a pan, and an espresso machine handle - placed on a cluttered table, as shown in Fig. 4.1. Initially, the hammer and pan were positioned in a pose where achieving high-quality grasps was not feasible, while the espresso machine handle was placed near an obstacle that made high-quality grasps unattainable due to collision. During running the experiment, we tasked the robot with moving those objects into a pre-placed bin using both the baseline method and grasp analyzer and multi-modal planner based method.

4.2.2 Results

Our multi-modal planner was successful in placing objects into the bin after a few trials. The baseline planner, on the other hand, was unable to successfully place the objects entirely. In the case of the hammer, once the gripper grasped the handle and lifted it up, the gravitational

Figure 4.1. Left: the setup of the real-world experiment with Fetch. Right: Objects (hammer, pan, and espresso machine handle) used in the real-world experiment.

torque caused the hammer to rotate about the contact point and remain in contact with the table 4.2. This caused the hammer to drag across the table during placement and ultimately grasp on the hammer failed. With our method, as shown in the lower four figures, our planner determined that the low-quality grasps on the handle were insufficient to lift the hammer, prompting the planner to rearrange the hammer to bring it closer. This adjustment enabled the robot to grasp the hammer in the middle where the predicted center of mass (COM) caused minimal gravitational torque and lift it up successfully.

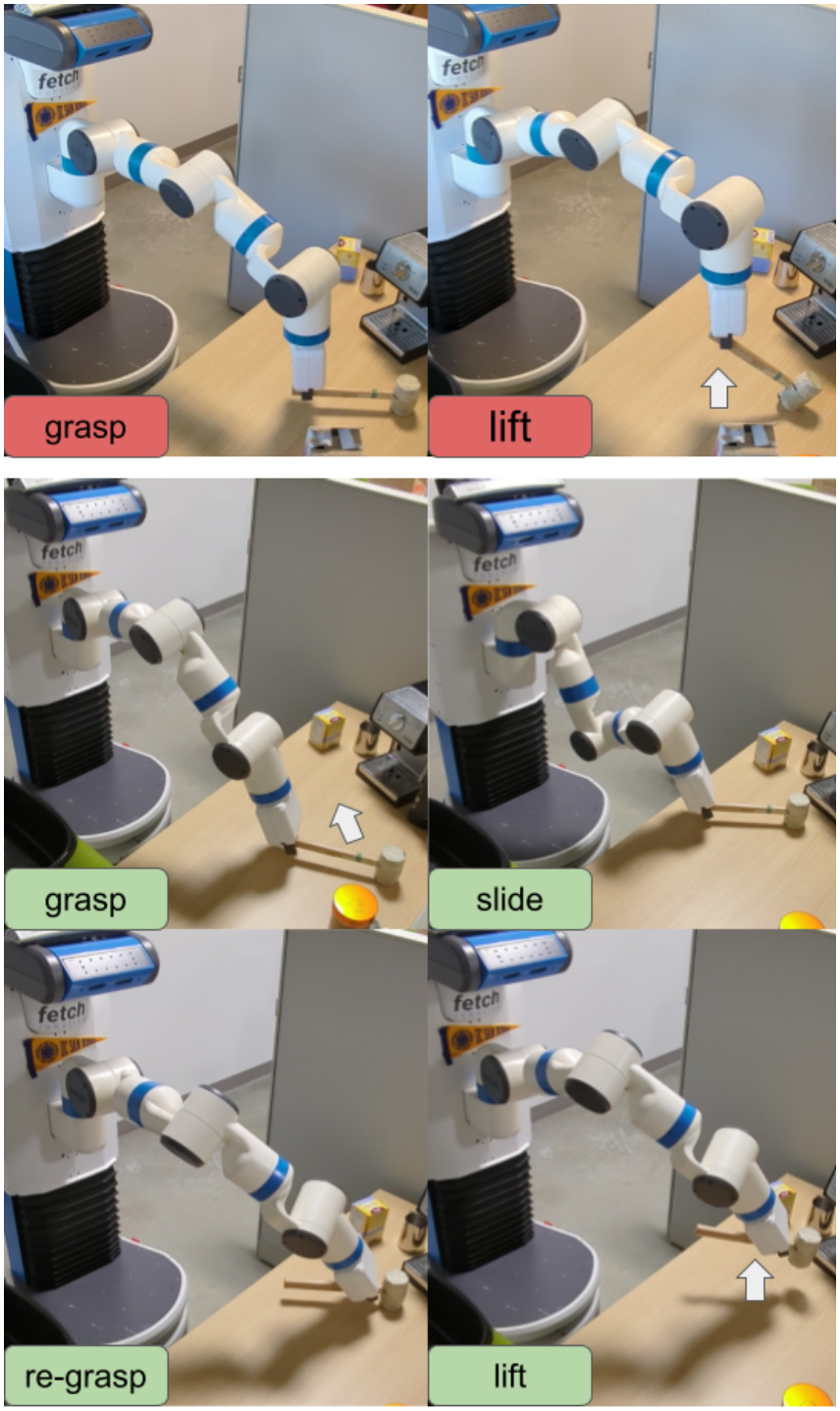


Figure 4.2. Upper: Without rearrangement, the hammer rotates around the grasping point during lifting. Lower: With rearrangement, before lifting, the robot identifies stable sliding and lifting grasp and slides the hammer into a better position for reliable lifting.

Chapter 5

Conclusions and Discussion

5.1 Conclusions

In this thesis we considered the problem of recommending grasps that enable stable manipulation of objects by a parallel gripper. We created a grasp analyzer that combines learning-based method with physics based methods that can identify stable sliding grasps and stable lifting grasps. The grasp analyzer works on unseen object with arbitrary geometric shapes. This grasp analyzer, combined with a multi-modal planner, was able to complete pick-and-place tasks in a stable manner both in-simulation and in the real world.

5.2 Discussion

One major weakness of the CoM predictor is objects with highly unevenly-distributed mass. This uneven distribution of mass usually originates from the use of different materials in different parts of the object. Therefore, a viable path for future work would be to add color as a feature onto the partial point cloud of the object. This new feature would allow the model to identify materials and thus predict uneven distribution in mass. This would require careful data generation or input processing as incorrect representation of material colors could lead to poor sim-to-real results.

Another potential direction for future work would be an end-to-end learning-based model that directly identifies stable sliding and lifting grasp poses. This could be achieved with a

distillation-like framework [6]. In such a setup, the output of the current learning-physics hybrid grasp analyzer would be used to generate learning targets for an end-to-end neural network. This could lead to improved efficiency and performance given that the network would directly predict stable sliding and lifting grasp poses.

Bibliography

[1] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository, 2015.

[2] Richard Cheng, Krishna Shankar, and Joel W. Burdick. Learning an optimal sampling distribution for efficient motion planning. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[3] B. Diziolu and K. Lakshiminarayana. Mechanics of form closure. *Acta Mechanica*, 52(12):107118, 1984.

[4] Clemens Eppner, Arsalan Mousavian, and Dieter Fox. Acronym: A large-scale grasp dataset based on simulation. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[5] Maximilian Haas-Heger. Grasp stability analysis with passive reactions. *CoRR*, abs/2103.06252, 2021.

[6] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network, 2015.

[7] Jiaming Hu, Zhao Tang, and Henrik I. Christensen. Multi-modal planning on rearrangement for stable manipulation.

[8] Dimitrios Kanoulas, Jinoh Lee, Darwin G. Caldwell, and Nikos G. Tsagarakis. Center-of-mass-based grasp pose adaptation using 3d range and force/torque sensing. *International Journal of Humanoid Robotics*, 15(04):1850013, Jan 2018.

[9] Arsalan Mousavian, Clemens Eppner, and Dieter Fox. 6-dof graspnet: Variational grasp generation for object manipulation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, May 2019.

[10] C. Qi, L. Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *NIPS*, 2017.

[11] Yuzhe Qin, Binghao Huang, Zhao-Heng Yin, Hao Su, and Xiaolong Wang. Dexpoint: Generalizable point cloud reinforcement learning for sim-to-real dexterous manipulation, 2022.

[12] E. Rohmer, S. P. N. Singh, and M. Freese. V-rep: A versatile and scalable robot simulation framework. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1321–1326, 2013.

[13] Martin Sundermeyer, Arsalan Mousavian, Rudolph Triebel, and Dieter Fox. Contact-graspnet: Efficient 6-dof grasp generation in cluttered scenes. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[14] Russ Tedrake. Robotic manipulation, Mar 2023.

[15] Andreas ten Pas, Marcus Gualtieri, Kate Saenko, and Robert Platt. Grasp pose detection in point clouds. *The International Journal of Robotics Research*, 36(1314):14551473, 2017.

[16] Chengde Wan, Thomas Probst, Luc Van Gool, and Angela Yao. Dense 3d regression for hand pose estimation, 2017.

[17] Jiangping Wang, Shirong Liu, Botao Zhang, and Qiang Lu. Motion planning with pose constraints based on direct projection for anthropomorphic manipulators. *IEEE Access*, 8:3251832530, Jan 2020.

[18] Melonee Wise, Michael Ferguson, Daniel King, Eric Diehr, and David Dymesich. Fetch & freight : Standard platforms for service robot applications. 2016.

[19] Yaxu Xue, Zhaojie Ju, Kui Xiang, Jing Chen, and Honghai Liu. Multiple sensors based hand motion recognition using adaptive directed acyclic graph. *Applied Sciences*, 7(4):358, Apr 2017.

[20] Yong Yu, K. Fukuda, and S. Tsujio. Estimation of mass and center of mass of graspless and shape-unknown object. *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, May 1999.