Tyler Franklin
12/25/2019

The Grasping Problem

Prof. James Meyer
MATH20300 Calculus III

The following aims to define and discuss the family of technical challenges known collectively as *the grasping problem*. First, some basic theory, notation, and definitions will be established. Next, the scope of the problem domain and its requirements will be described, illustrating technical difficulties and limitations. Finally, future work in specific technical areas will be proposed, with some discussion of potential benefits to industry.

I. Introduction

Picking up and manipulating arbitrary objects is a deceptively complex task, yet the average human masters it as a child and uses the skill daily without much thought. For this reason, it may come as a surprise that it represents a famously intractable problem in the fields of robotics, computer vision, and machine learning. While we have made a lot of progress, the problem remains unsolved insofar as parity with human ability is concerned.[1]

Most of the complexity of the problem is a direct result of variability in task parameters.[2] That is, if all parameters are fixed, such as on a factory assembly line, a repetitive grasping routine is commonplace and can be configured for precision and reliability. When variability is introduced to the program, however, choosing a complementary internal state requires some understanding of the external state, a significantly more challenging task.[3]

II. Basic Theory

We will assume Coulomb friction wherein fixed contact is maintained for normal contact force F_i^n and tangent contact force F_i^t, which include external forces, at any given hard-finger contact point c_i if and only if $|F_i^t| \leq \mu_s F_i^n$, outside of which the contact type switches from fixed to sliding.[4] This describes a symbolic friction cone C whose angle $\theta = \tan^{-1}\mu$ (fig. 1), i.e. acute or obtuse for lower or higher μ_s, respectively.

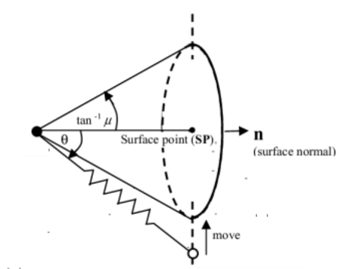

Figure 1. *Typical illustration of a friction cone.*[5]

In short, this means that F_s exists at surface contact point c_i for friction cone C if $c_i \in C$ is true. Outside of this contact stability, $\vec{v} > 0$ and grip is governed by F_k which we consider a failure condition.[6]

Each wrench $w_i \in W$ is a description of holonomic force effects on c_i that includes three-dimensional force and torque components with respect to some R.[7] That is, $w_i = (\ldots F_i, \ldots \tau_i)$ where $F_i = \mu_i F_i^n = (f_x, f_y, f_z)$ translational force components and $\tau_i = (c_i - R) \times F_i = (\tau_x, \tau_y, \tau_z)$ rotational force components.[8] This means that maintaining equilibrium at c_i requires that w_i cancel any external wrench w_{ext}, like gravity or environmental turbulence, such that $w_i + w_{ext} = \vec{0}$. Likewise, maintaining equilibrium for any force-closed grasp means, simply:

$$\sum_{i=1}^{N} w_i + w_{ext} = \vec{0}$$

Force-closure is a property of a grasp for which it is feasible to counteract w_{ext} at some point with corresponding $w_i \ldots w_n$ at some $c_i \ldots c_n$.[9] In order for this to be true, all potential forces in six degrees must be covered.

Thus, a common test is to check if the origin R, shared by $c_i \ldots c_n$ lies within an estimated grasp wrench space (GWS).[10] This space spans the projective vectors at the boundaries of each friction cone.[11] For a single contact point c_i this would simply be the cone itself, but for two or more friction cones, the GWS forms an internal subspace of the object:

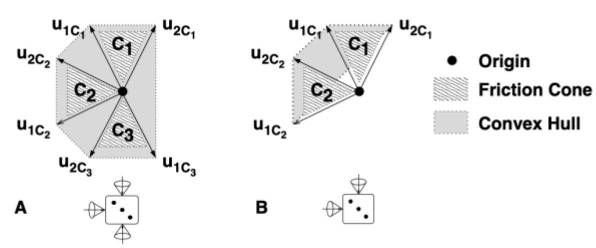

Figure 1. *A grasp wrench space in 2D. A contains the origin. B does not.*[12]

The notion of a GWS has further utility. In the definition of F_i above, we rely on μ_i which, given our understanding of the friction cone, will vary according to the angle of force, with an F_i parallel to the surface normal representing $\mu_s F$ exactly, and any other angle representing some lesser force. An inexpensive

[1] Christian Ernst, Joachim Rupprecht, "Learning under Ambiguity through Multiple Predictions " (PhD diss., Technische Universität München, 2018), https://d-nb.info/119244177X/34p, 53.
[2] Richard Hodson, "How Robots Are Grasping the Art of Gripping," May 9, 2018, https://www.nature.com/articles/d41586-018-05093-1, para. 3.
[3] Ibid., para. 5.
[4] Stephane Caron, Quang-Cuong Pham, and Yoshihiko Nakamura, "Stability of Surface Contacts for Humanoid Robots: Closed-Form Formulae of the Contact Wrench Cone for Rectangular Support Areas," *2015 IEEE International Conference on Robotics and Automation (ICRA),* 2015, https://doi.org/10.1109/icra.2015.7139910, 4-5.
[5] N. Melder and W. S. Harwin, "Extending the Friction Cone Algorithm for Arbitrary Polygon Based Haptic Objects," *12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS 04. Proceedings,* 2004, https://doi.org/10.1109/haptic.2004.1287201, 2.
[6] This is not necessary, strictly speaking. For an example of responsive force management using slip awareness, see Morita Nobutomo, Hirofumi Nogami, Eiji Higurashi, and Renshi Sawada, "Grasping Force Control for a Robotic Hand by Slip Detection Using Developed Micro Laser Doppler Velocimeter," *Sensors* 18, no. 2 (2018): 326, https://doi.org/10.3390/s18020326.
[7] C. Borst, M. Fischer, and G. Hirzinger, "Grasping the dice by dicing the grasp," *IEEE/RSJ IROS*, 2003, 3693. This means each wrench is always considered with respect to some reference point, such as the center of mass. This becomes important in the next section.
[8] Ibid.
[9] Ibid., 3694.
[10] Ibid.
[11] Ibid.
[12] N. Melder and W. S. Harwin, "Extending the Friction Cone Algorithm for Arbitrary Polygon Based Haptic Objects," *12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS 04. Proceedings,* 2004, https://doi.org/10.1109/haptic.2004.1287201, 2.

solution is offered by Pollard.[13] Common gripping hardware with coplanar contact points, such as those styled after the Salisbury hand, will reduce the GWS to a plane.[14] In this case, any viable grasp plane at c_i will slice the friction cone at some angle θ_i (fig. 2).

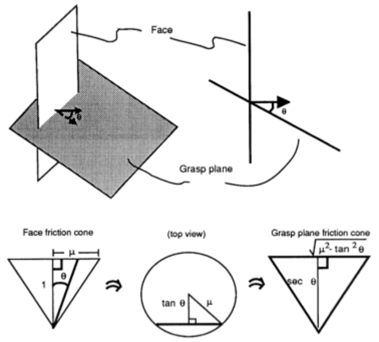

Figure 2. *A grasp plane slicing the friction cone of a given face.*[15]

This allows us to calculate an effective friction coefficient μ_i at c_i, based on the angle θ_i between the surface normal $\bar{n}$ at c_i and the contact force F_i:[16]

$$\mu_i = \frac{\sqrt{\mu_s^2 - \tan^2\theta_i}}{\sec\theta_i}$$

Methods such as this can simplify the optimization problem of choosing the grasping points to maintain force closure and maximize grip.

III. Problem Scope

We have established that a secure, force-closed grasp is one that exhibits equilibrium of forces on an object throughout manipulation such that $\sum_{i=1}^{N} w_i + w_{ext} = \vec{0}$. To achieve this, as established above, the grasping routine must counteract w_{ext}, accounting for gravity and marginal forces in 6D from inertia and external turbulence at any given point, with $w_i \ldots w_n$ for every point of contact $c_i \ldots c_n$.[17]

Here, the hypothetical grasping routine runs into additional data requirements immediately, since $F_i = \mu_i F_i^n$ and μ_i as defined above requires μ_s. We know that the coefficient μ_s is unique to the contact surfaces' materials and must either be known in advance, determined given sensor inputs, or replaced with a static value for all surface types, to know what value of F_n is required.[18]

Similarly, since a given factor of w_{ext} is gravity $F_g = mg$, to counteract this force the program must determine, or know in advance, the mass m of the object to accurately determine $w_i \ldots w_n$. A cheap and obvious fix here might be to overestimate, such that any object up to some predetermined threshold could be hefted. This is arguably another problematic strategy, however, since many lighter objects cannot withstand excess force.[19]

In addition, a stable configuration of $c_i \ldots c_n$ securing an object is difficult to guarantee without some knowledge of its shape.[20] For example, Pollard uses three surface normals of a given polyhedral to restrict the range of the viable grasp planes:[21]

$$N = \begin{bmatrix} n_{1x} & n_{1y} & n_{1z} \\ n_{2x} & n_{2y} & n_{2z} \\ n_{3x} & n_{3y} & n_{3z} \end{bmatrix}$$

Where $rank[N]$ determines the possible orientations of the grasp plane such that $rank[N] \in \{1,2\}$, two or more faces have parallel normals, allows a perpendicular grasp plane with maximum μ_s and $rank[N] = 3$ are linearly independent normal which require solving:[22]

$$N\vec{g} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Where $\overrightarrow{g^T}$ is the normal of the grasp plane.[23] In this case the grasp plane here will intersect all the faces $c_i \ldots c_n$ at equal angles $\theta_i \ldots \theta_n$, such that effective friction coefficient $\mu_i \ldots \mu_n$ is equal.

This falls squarely in the realm of computer vision, as interpreting the dimensions of real-world objects with clarity and precision, in real-time, is an ongoing pursuit in the field.[24]

Another information deficit related to shape is density mapping. This concerns our definition of wrench points $w_i \ldots w_N$ where $\tau_i = (c_i - R) \times F_i$, since R is an arbitrarily assigned point. Many everyday objects have graspable surfaces which humans intuitively select based on anticipated mass distribution, such that $\sum_{i=1}^{N} c_i - R$ is minimized.[25] This relates directly to force requirements since any grasping configuration

[13] Nancy S. Pollard, "The Grasping Problem: Toward Task-Level Programming for an Articulated Hand." Boston, MA: MIT CSAIL, 1990. https://doi.org/http://hdl.handle.net/1721.1/6817, 25.
[14] C. Pellerin, "Salisbury Hand", *Industrial Robot*, vol. 18 no. 4 (1991), https://doi.org/10.1108/eb004571, 25-6.
[15] Pollard, 26.
[16] Ibid., 25.
[17] Morita, 1. "Effective" used here means with consideration of the limiting nature of the surface with coefficient μ_s. Also, while ventral F_n such as that of a forklift is indeed possible in practice, we ignore it here as outside the scope as part of form-closure and caging strategies.
[18] Kaiyu Hang, Florian T. Pokorny, and Danica Kragic, "Friction Coefficients and Grasp Synthesis," *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, November 3, 2013, https://doi.org/10.1109/iros.2013.6696858, 1. One can imagine that the latter is potentially problematic for slippery, dense objects, like ice cubes or wet soap bars, and grippy, brittle objects such as dried flowers or saltine crackers.
[19] For example, the force required to lift a ceramic plate out of a dishwasher might crush a paper cup.
[20] Ashutosh Saxena, Justin Driemeyer, and Andrew Y. Ng, "Robotic Grasping of Novel Objects Using Vision," *The International Journal of Robotics Research* 27, no. 2 (2008), https://doi.org/10.1177/0278364907087172, 157.
[21] Pollard, 25.
[22] Ibid., 28.
[23] Ibid.
[24] Derek Hoiem, and Silvio Savarese, "Representations and Techniques for 3D Object Recognition and Scene Interpretation," *Synthesis Lectures on Artificial Intelligence and Machine Learning 5*, no. 5 (2011), 1.
[25] By way of example, consider the placement of handles on objects like luggage and boxes. These are usually positioned with some consideration of anticipated center of mass. Clearly, objects *without* handles still have these optimal grasping points. Arguably, humans often use these points instinctively.

without redundancy will be limited by the weakest w_i where $w_{ext} > w_i$.[26]

In summary, the key areas of difficulty are programmatic understanding of surface material, object mass, object shape, and mass distribution. From the perspective of any given grasping procedure, these are all crucial variables. Conventionally, each of these might be characterized as exclusively computer vision problems, however we will argue that these areas also meet at the technical challenge of physical simulation.

IV. Future Work

While there are likely numerous potential avenues for enabling more robust solutions to the grasping problem, it seems the two technical areas with the most potential leverage are vision and physical simulation. Both strike at the information deficits described above and are currently fairly active research spaces.

In regards to vision, a strategy that might be worth exploring further is cognitive context.[27] This seeks to understand and emulate the human ability to quickly recognize patterns via contextual associations, pursuing the "cognitive computing" aspect of data science.[28]

Recent advances have been made in the area of machine learning, not just for the vision aspect of the problem, but also the geometric computation. Some strategies use imitation learning, training networks with data from human agents.[29] Others use synthetic world modeling to quickly train networks.[30] Still others outsource the entirety of the problem, including vision and geometry, to the proverbial black-box of a convolutional neural network, and let it work out the rules for itself.[31]

While it would seem the scalability cost of using physical hardware for real-time training might make it a less attractive option, it skips the problem of physical simulation entirely. Also, its verifiable success might highlight deficiencies in our ability to lend the skill of pre-visualization to robotics in the way of physical simulation. If we are able to more easily and accurately model the physical world, we may be able to better predict the data missing from the geometric side of the problem.

Solving the grasping problem, even for limited cases, has immediate implications for industrial and logistics applications, such as warehouse management for large distributors like Amazon, who invest heavily in logistic automation.[32] For example, Amazon holds an annual "picking challenge" for pick-and-place technology.[33] The tests involve grasping miscellaneous target objects from bins filled with common, miscellaneous Amazon items, quite similar to the item-bin training trials we saw in Levine, et al.[34]

Similarly, fast food chains like McDonalds have recently implemented automative techniques related to their food ordering process, but there lack parallel solutions for their food preparation process. One can imagine a holonomic robot with human like dexterity could easily be trained to execute food orders for some portion of the menu.

Human-like dexterity, however, opens doors in far more sectors than just industrial and logistics or food service.[35] Further applications might include medicine, automotive repair, and domestic trade work, to name a few. In addition, any physical task that requires a high level of precision, speed, or safety, is one that might be better delegated to the reliability of automation.

IV. Conclusion

It is, perhaps, not too difficult to imagine how lucrative workforce automation can be for large companies like Amazon or McDonalds, especially in settings requiring strict safety and/or health protocols.[36] This may lead to a concern that advancement of technology like this will take jobs from the human workforce. While this may technically be possible, human-like grasping also has implications as assistive technology, aiding employees in their jobs, and enabling the disabled in daily activities.[37]

Ultimately solving the grasping problem in its entirety, including the technical limitations outlined above, will likely not occur all at once. It is far more likely that we adopt increasingly expedient strategies for solving an ever expanding cross section of proposed use cases, as we have to date.

[26] Pollard, 25.

[27] Lawson Wallace, Laura Hiatt, and J. Gregory Trafton, "Leveraging Cognitive Context for Object Recognition," *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops,* 2014, https://doi.org/10.1109/cvprw.2014.63.

[28] Jun Wu, "AI and Cognitive Computing," Medium, Towards Data Science, July 12, 2019, https://towardsdatascience.com/ai-and-cognitive-computing-fc701b4fbae7.

[29] Tianhao Zhang, Zoe Mccarthy, Owen Jow, Dennis Lee, Xi Chen, Ken Goldberg, and Pieter Abbeel, "Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018. https://doi.org/10.1109/icra.2018.8461249, et al.

[30] Saxena, 4-5. One interesting observation made in this study was that the quality of generated images was positively correlated with trial success up to the point where the rendering engine became a limiting factor.

[31] Sergey Levine, Peter Pastor, Alex Krizhevsky, Julian Ibarz, and Deirdre Quillen, "Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection," *The International Journal of Robotics Research* 37, no. 4-5 (2017): 421–36. https://doi.org/10.1177/0278364917710318, et al. It appears this study marks a recent trend of generating robotics training data in real-time with physical hardware. See

[32] Daniel Wahrmann, Arne-Christoph Hildebrandt, Christoph Schuetz, Robert Wittmann, and Daniel Rixen, "An Autonomous and Flexible Robotic Framework for Logistics Applications," *Journal of Intelligent & Robotic Systems* 93, no. 3-4 (May 2017): 419–31 https://doi.org/10.1007/s10846-017-0746-8, 428-30.

[33] Peter R. Wurman, and Joseph M. Romano, "The Amazon Picking Challenge 2015 [Competitions]," *IEEE Robotics & Automation Magazine* 22, no. 3 (2015), https://doi.org/10.1109/mra.2015.2452071, 10.

[34] Ibid.

[35] [1] Correll, Niklaus. "Robots Getting a Grip on General Manipulation." IEEE Spectrum. November 13, 2018. https://spectrum.ieee.org/automaton/robotics/industrial-robots/robots-getting-a-grip-on-general-manipulation

[36] Will Evans, "Ruthless Quotas at Amazon Are Maiming Employees," *The Atlantic,* Atlantic Media Company, December 5, 2019, https://www.theatlantic.com/technology/archive/2019/11/amazon-warehouse-reports-show-worker-injuries/602530/.

[37] Andrés Úbeda, Brayan S. Zapata-Impata, Santiago T. Puente, Pablo Gil, Francisco Candelas, and Fernando Torres, "A Vision-Driven Collaborative Robotic Grasping System Tele-Operated by Surface Electromyography," *Sensors* 18, no. 7 (2018), https://doi.org/10.3390/s18072366, 2366.

References

[1] Bologni, L. "Robotic Grasping: How To Determine Contact Positions." *Robot Control 1988 (Syroco 88)*, 1989, 395–400. https://doi.org/10.1016/b978-0-08-035742-3.50070-8.

[2] Borst, C., M. Fischer, and G. Hirzinger, "Grasping the dice by dicing the grasp," *IEEE/RSJ IROS*, 2003, 3692–3697.

[3] Caron, Stephane, Quang-Cuong Pham, and Yoshihiko Nakamura. "Stability of Surface Contacts for Humanoid Robots: Closed-Form Formulae of the Contact Wrench Cone for Rectangular Support Areas." *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015. https://doi.org/10.1109/icra.2015.7139910.

[4] Chinellato, Eris, and Del Pobil Angel P. The Visual Neuroscience of Robotic Grasping: Achieving Sensorimotor Skills through Dorsal-Ventral Stream Integration. Cham: Springer, 2015.

[5] Correll, Niklaus. "Robots Getting a Grip on General Manipulation." IEEE Spectrum. November 13, 2018. https://spectrum.ieee.org/automaton/robotics/industrial-robots/robots-getting-a-grip-on-general-manipulation

[6] Domae, Yukiyasu. "Amazon Picking Challenge 2016." *Journal of the Robotics Society of Japan* 34, no. 10 (2016): 696–96. https://doi.org/10.7210/jrsj.34.696.

[7] Ernst, Christian, Joachim Rupprecht. "Learning under Ambiguity through Multiple Predictions." PhD diss., Technische Universität München, 2018. https://d-nb.info/119244177X/34

[8] Evans, Will. "Ruthless Quotas at Amazon Are Maiming Employees." *The Atlantic*. Atlantic Media Company, December 5, 2019. https://www.theatlantic.com/technology/archive/2019/11/amazon-warehouse-reports-show-worker-injuries/602530/.

[9] Hang, Kaiyu, Florian T. Pokorny, and Danica Kragic. "Friction Coefficients and Grasp Synthesis." *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, November 3, 2013. https://doi.org/10.1109/iros.2013.6696858.

[10] Hodson, Richard. "How Robots Are Grasping the Art of Gripping," May 9, 2018. https://www.nature.com/articles/d41586-018-05093-1.

[11] Hoiem, Derek, and Silvio Savarese. "Representations and Techniques for 3D Object Recognition and Scene Interpretation." *Synthesis Lectures on Artificial Intelligence and Machine Learning 5*, no. 5 (2011): 1–169. https://doi.org/10.2200/s00370ed1v01y201107aim015.

[12] Lawson, Wallace, Laura Hiatt, and J. Gregory Trafton. "Leveraging Cognitive Context for Object Recognition." *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014. https://doi.org/10.1109/cvprw.2014.63.

[13] Lenz, Ian, Honglak Lee, and Ashutosh Saxena. "Deep Learning for Detecting Robotic Grasps." *Robotics: Science and Systems IX*, 2013. https://doi.org/10.15607/rss.2013.ix.012.

[14] Levine, Sergey, Peter Pastor, Alex Krizhevsky, Julian Ibarz, and Deirdre Quillen. "Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection." *The International Journal of Robotics Research* 37, no. 4-5 (2017): 421–36. https://doi.org/10.1177/0278364917710318.

[15] Melder, N., and W.s. Harwin. "Extending the Friction Cone Algorithm for Arbitrary Polygon Based Haptic Objects." *12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS 04. Proceedings.,* 2004. https://doi.org/10.1109/haptic.2004.1287201.

[16] Morita, Nobutomo, Hirofumi Nogami, Eiji Higurashi, and Renshi Sawada. "Grasping Force Control for a Robotic Hand by Slip Detection Using Developed Micro Laser Doppler Velocimeter." *Sensors* 18, no. 2 (2018): 326. https://doi.org/10.3390/s18020326.

[17] Pellerin, C., "Salisbury Hand", *Industrial Robot*, Vol. 18 no. 4 (1991): 25-6. https://doi.org/10.1108/eb004571

[18] Pollard, Nancy S. "The Grasping Problem: Toward Task-Level Programming for an Articulated Hand." *The Grasping Problem: Toward Task-Level Programming for an Articulated Hand*. Boston, MA: MIT CSAIL, 1990. https://doi.org/http://hdl.handle.net/1721.1/6817.

[19] Salisbury, J. K., and B. Roth. "Kinematic and Force Analysis of Articulated Mechanical Hands." *Journal of Mechanisms, Transmissions, and Automation in Design* 105, no. 1 (January 1983): 35–41. https://doi.org/10.1115/1.3267342.

[20] Saxena, Ashutosh, Justin Driemeyer, and Andrew Y. Ng. "Robotic Grasping of Novel Objects Using Vision." *The International Journal of Robotics Research* 27, no. 2 (2008): 157–73. https://doi.org/10.1177/0278364907087172.

[21] Úbeda, Andrés, Brayan S. Zapata-Impata, Santiago T. Puente, Pablo Gil, Francisco Candelas, and Fernando Torres. "A Vision-Driven Collaborative Robotic Grasping System Tele-Operated by Surface Electromyography." *Sensors* 18, no. 7 (2018): 2366. https://doi.org/10.3390/s18072366.

[22] Vezzani, Giulia, and Lorenzo Natale. "Real-Time Pipeline for Object Modeling and Grasping Pose Selection via Superquadric Functions." *Frontiers in Robotics and AI* 4 (2017). https://doi.org/10.3389/frobt.2017.00059.

[23] Wahrmann, Daniel, Arne-Christoph Hildebrandt, Christoph Schuetz, Robert Wittmann, and Daniel Rixen. "An Autonomous and Flexible Robotic Framework for Logistics Applications." *Journal of Intelligent & Robotic Systems* 93, no. 3-4 (May 2017): 419–31. https://doi.org/10.1007/s10846-017-0746-8.

[24] Wu, Jun. "AI and Cognitive Computing." Medium. Towards Data Science, July 12, 2019. https://towardsdatascience.com/ai-and-cognitive-computing-fc701b4fbae7.

[25] Wurman, Peter R., and Joseph M. Romano. "The Amazon Picking Challenge 2015 [Competitions]." *IEEE Robotics & Automation Magazine* 22, no. 3 (2015): 10–12. https://doi.org/10.1109/mra.2015.2452071.

[26] Zhang, Tianhao, Zoe Mccarthy, Owen Jow, Dennis Lee, Xi Chen, Ken Goldberg, and Pieter Abbeel. "Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation." *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018. https://doi.org/10.1109/icra.2018.8461249.