Student Responses to RENEWAL Questions

The following apply only for renewal applications. Please consider these responses in making your assessment of the application.

1. Please describe how successful you were in achieving the intended outcomes of and adhering to the plan/timeline of your original proposal.

Our original point introduced two main focuses: incorporating collision constraints into grasping and improving 3D comprehension amidst clutter. After further investigation, Prof. [redacted] and I decided the first point alone, collision constraints for grasping, warranted a full summer of attention. We followed the proposed schedule well, spending the first month formulating our approach and the subsequent 2 months implementing and testing. Our work successfully culminated in a submission to the IEEE International Conference on Robotics and Automation (ICRA) 2020, currently under review.

2. Please describe how successful you think your relationship with your mentor was during your first semester of UROP.

Professor [redacted] and I have a strong relationship. Prof. [redacted] helped me formulate and setup my problem statement with regards to collision constraints for grasping and provide many pointers to current literature in the community related to my project. Throughout development, we had weekly meetings specific to my project where he provided both high and low-level guidance as needed. He also helped facilitate help from other lab members as needed and communicated with other contacts outside the university on the project to provide me with further feedback and guidance.

3. Please explain what you think the impact of an additional semester in UROP would be to your educational and career goals.

My first semester with UROP brought me into a research field I am now interested in turning into a full graduate career. My work exploring uniting perception and action through my UROP Summer 2019 project inspired my current research work on scene reconstruction I am doing as a research intern at another university and formed the basis of my NSF Graduate Fellowship proposal. Another semester will give me the opportunity to continue exploring the space, refining my interests, and allowing me to continue to contribute to an area of research I am very excited by.

UROP Proposal

Title of Proposal

Robust Object Reconstruction for Robotic Grasping

Problem/Topic of Research or Creative Work

Generalizable grasping, in which a robotic arm can grasp and lift a large variety of both previously seen and unseen objects in multiple environments, remains an active and important research topic in robotics. The problem is made difficult in part by only having access to noisy and partial sensor readings of the scene. Uncertainty surrounding object geometries and dynamic properties makes it difficult to effectively synthesize grasps for unknown objects. To overcome this, recent successes in robotics research have relied on deep learning techniques, which are capable of effectively handling the noise and complexity inherent to the problem.

Most existing deep learning grasp synthesis approaches train in an end-to-end fashion (e.g., Lu et al. 2019, Mahler et al. 2019), that is, they act directly on the sensor input to perform grasp synthesis. This makes the assumption that the deep learning framework can implicitly learn a geometric understanding. We hypothesize that explicitly reasoning over the geometry of the scene will yield better grasp synthesis results. Specifically, we believe the subtask of identifying full object geometry, reconstructed from the noisy partial sensor readings, is vital to grasping success. This intuition inspired our Summer 2019 UROP project, in which we created an object reconstruction algorithm, PointSDF, and used it to improve robotic grasping (Van der Merwe et al.).

While the object reconstruction learned by PointSDF enabled improved grasp synthesis, our approach currently has several shortcomings. First, it does not model uncertainty in reconstruction in a principled fashion.

As such, we are unable to leverage the active nature of robots to intelligently gather more information to improve our object understanding (Bajcsy et. al., Bohg et. al. 2017), nor incorporate uncertainty into the grasp selection (Lundell et al., Mahler et al. 2015, Li et al.). We propose to extend our PointSDF representation to address these shortcomings, yielding robust object reconstructions for robotic grasping.

First, we propose to extend PointSDF to be uncertainty-aware (Gal et al., Kendall et al.). This lets the reconstruction capture ambiguities in sensor occlusion, e.g., a mug handle could have many different shapes if not viewed directly, as well as prevent ill-informed grasps, allowing the robot to continue sensing until it is confident or ask for human supervision before grasping. Next, I will extend PointSDF to utilize active perception (Bajcsy et al., Bohg et al. 2017) via multi-modal inputs. Using the uncertainty awareness, the robot can select an area of the object for which it is most uncertain, then either a) view it using the camera, or b) reach out and touch with a tactile sensor, depending upon the location of interest and the environment. This disambiguates the reconstruction while handling failure cases of each sensing modality, such as limited reach, control error, and obstructions in line of sight.

We then propose to utilize the reconstruction to plan grasps on a real-robot system. I propose to explore two grasp planning approaches. First, I will use analytical grasp geometry metrics given the predicted object geometry. Second, I will use a deep learning model to learn a grasp metric directly, extending our grasping model from our earlier UROP project (Van der Merwe et al.).

Relevant Background/Literature Review

While robotic grasping has been studied extensively in the robotics research community, it remains an open and difficult research question, especially for multi-fingered hands. Approaches to robotic grasping can be split into two main approaches: analytical, in which the grasp is synthesized using the dynamics of the object and grasp, and data-driven, in which the grasp is synthesized using a learning based approach (Sahbani et al., Bohg et al. 2013). While analytical approaches have strong guarantees, they tend to be computationally expensive and rely on significant knowledge of the object to be grasped. Data-driven approaches have shown state of the art results, capable of handling the complexity of grasping (Mahler et al. 2019, Lu et al. 2017, Kappler et al.).

Current deep learning approaches take in sensor input, such as RGB or RGBD image, and output a grasp, either via direct regression (Liu et al., Veres et. al), sampling candidate grasps or motions (Mahler et al. 2019, Levine et al.), or solving an optimization problem leveraging the learned network, either in a discrete (Zeng et al, Pinto et al.) or continuous (Lu et al. 2018, Lu et al. 2019, Zhou et al., Varley et al. 2015) fashion. We propose to split grasping into the subtask of deriving object geometry via reconstruction first, followed by a continuous optimization problem utilizing that information through collision constraints (Lu et al. 2018, Van der Merwe et al.). Few approaches have similarly split the grasping task and existing approaches that do rely on comparatively low resolution object reconstructions (Varley et al. 2017, Yan et al.).

Learning object representations and reconstructions has long been a focus in the computer vision community, and multiple representations have been utilized in reconstructions, including voxel-based (Brock et al., Choy et al.) and point-cloud based (Fan et al.) approaches. Recent state of the art results in reconstruction seek instead to learn an implicit boundary, in which arbitrary query points in 3D space are assigned either occupancy (Mescheder et al., Chen et al. 2018) or signed distance (Park et al.) labels with regards to the represented object. Our PointSDF object modeling approach is based closely on these recent works (Mescheder et al., Park et al.).

While deep learning results are powerful, most modern approaches do not directly model the uncertainty of the predictions and network, increasing the risk of wrong predictions leading to dangerous outcomes. Recent work has explored bringing explicit uncertainty modeling into deep neural networks by using probabilistic loss functions and applying sampling techniques to predictions (Gal et al., Kendall et al.). Most modern deep learning approaches also rely solely upon a single sense reading from the object. Active and Interactive perception are a form of sensing that use the active nature of the robot to collect more information about the target environment before attempting a final task (Bajcsy et al., Bohg et al. 2017). Uncertainty can be used to guide active perception, collecting more information to reduce the uncertainty in model predictions. This idea has been applied to 3D reconstruction, but the approach relies on limited uncertainty quantification, a low-resolution representation, and limit active exploration only to touch (Wang et al.). Our proposal enables higher resolution, more modalities, and more accurate uncertainty estimation than existing approaches and applies the representation to improve grasping.

Specific Activities to be Undertaken and Timeframe for Each Activity

1. November-December: Literature Review + Baselines:
Before the funding period, I will be completing a research internship at Mila (Montreal Institute of Learning Algorithms) where I am applying implicit surface reconstruction techniques to scene comprehension. This involves coding approaches that will become direct baselines to our proposed approach. During this time I will also continue to explore the existing literature on uncertainty in representation, active perception based on uncertainty, and robotic grasping.
2. January 6 - February 1: Uncertainty in 3D Reconstruction:
Upon completion of the baseline methods, and during the start of the funding period, we propose to formulate and implement uncertainty in our 3D reconstruction approach. For uncertainty based representations, we will extend our current PointSDF network architecture to include a probabilistic loss and dropout sampling to reflect uncertainty in reconstructions (Kendall et al., Gal et al.). We will then collect training data with which to train our network, which involves using a simulated rendering pipeline created during our Summer 2019 UROP project. We will then train and validate our network on our data and benchmark to baselines that do not incorporate uncertainty (Mescheder et al., Park et al.). We will also apply our approach qualitatively on real world data, collected with a Kinect sensor, to ensure its viability for use with a real robot system. As much of the baselines will be in place prior to the funding period, we believe we can extend and complete this task in the first month.
3. February 1-April 1: Active Perception in 3D Reconstruction:
To perform active perception, we must first determine the location to sense. We propose to solve this by searching the surface of the predicted object reconstruction for the location of highest uncertainty, via an optimization problem. At this point in the project, we will benchmark several approaches for solving this optimization. We will then need to formulate an algorithm for turning the uncertainty point into a viewpoint for the camera, by utilizing the predicted surface normal and back projecting into free space. Once done, we will implement the system to execute the plan to observe the uncertain point, first in simulation with simulated sensors, then on the real system. This procedure has many components, and thus we give the bulk of the funding period to this development time.
4. April 1- April 29: Grasping via 3D Reconstruction:
Our last month of the funding period is dedicated to grasping. We will implement a geometrically grounded analytic metric as a baseline approach (Sahbani et al.) as well as implement an extension to our proposed deep learning grasp metric from our previous UROP project (Van der Merwe et al.). We will then formulate and implement an optimization over the grasp metric. This will involve exploring several optimization techniques in order to make the optimization efficient. Finally we will deploy the grasp planning approach on simulation and real-robot grasping benchmarks.
5. Post-Funding Period: Writing, Submission:
After the funding period, we will continue our writeup and finish testing our system. We then plan to submit our work to a peer-reviewed conference, specifically the Conference on Robotic Learning (CoRL) 2020.

Relationship of the Proposed Work to the Expertise of the Faculty Mentor

Professor [redacted] is an Assistant Professor at the School of Computing, and is the head of the Utah Learning Lab for Manipulation Autonomy. The lab and Professor [redacted] are very active in manipulation research, and have published multiple recent papers on multi-fingered grasping, of which this project is an extension (Lu et al. 2018, Lu et al. 2019). I spent the Summer of 2019 as a UROP student in Professor [redacted]'s lab, where I interacted daily with graduates students actively researching in this area as well as get advising from Professor [redacted] during 1-on-1 and group meetings. Our work culminated in a conference submission currently under review (Van der Merwe et al.).

Relationship of the Proposed Work to Student's Future Goals

I plan to pursue research long-term in robotics and computer vision work, including a doctorate on this subject. I am especially interested in the intersection of computer vision with embodied agents (such as robots and self-driving vehicles) and in bridging the gap between action and perception. This work allows me to explore cutting edge work in both computer vision and robotics, as well as learning new skills in motion planning, optimization, and deep learning. This work will form the basis of my graduate career, which I plan to begin in the Fall of 2020. Under Professor [redacted]'s leadership, I have already gained engineering and research skills, as well as learning to collaborate in a research environment, and I would like to continue to develop those skills as I look forward to a career in the research community.

References

Mahler, Jeffrey, et al. "Learning ambidextrous robot grasping policies." Science Robotics 4.26 (2019).

Lu, Qingkai, et al. "Planning Multi-Fingered Grasps as Probabilistic Inference in a Learned Deep Network." International Symposium on Robotics Research. 2017.

Kappler, Daniel, Jeannette Bohg, and Stefan Schaal. "Leveraging big data for grasp planning." 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2015.

Zeng, Andy, et al. "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching." 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018.

Pinto, Lerrel, and Abhinav Gupta. "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours." 2016 IEEE international conference on robotics and automation (ICRA). IEEE, 2016.

Levine, Sergey, et al. "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection." The International Journal of Robotics Research37.4-5 (2018): 421-436.

Varley, Jacob, et al. "Generating multi-fingered robotic grasps via deep learning." 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2015.

Lu, Qingkai, and Tucker Hermans. "Modeling Grasp Type Improves Learning-Based Grasp Planning." IEEE Robotics and Automation Letters 4.2 (2019): 784-791.

Zhou, Yilun, and Kris Hauser. "6DOF Grasp Planning by Optimizing a Deep Learning Scoring Function." Robotics: Science and Systems (RSS) Workshop on Revisiting Contact-Turning a Problem into a Solution. 2017.

Liu, Min, et al. "Generating Grasp Poses for a High-DOF Gripper Using Neural Networks." arXiv preprint arXiv:1903.00425 (2019).

Veres, Matthew, Medhat Moussa, and Graham W. Taylor. "Modeling grasp motor imagery through deep conditional generative models." IEEE Robotics and Automation Letters2.2 (2017): 757-764.
Lundell, Jens, Francesco Verdoja, and Ville Kyrki. "Robust Grasp Planning Over Uncertain Shape Completions." arXiv preprint arXiv:1903.00645 (2019).

Mahler, Jeffrey, et al. "Gp-gpis-opt: Grasp planning with shape uncertainty using gaussian process implicit surfaces and sequential convex programming." 2015 IEEE international conference on robotics and automation (ICRA). IEEE, 2015.

Li, Miao, et al. "Dexterous grasping under shape uncertainty." Robotics and Autonomous Systems 75 (2016):

352-364.

Gal, Yarin, and Zoubin Ghahramani. "Dropout as a bayesian approximation: Representing model uncertainty in deep learning." international conference on machine learning. 2016.

Joon Park, Jeong, et al. "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

Sahbani, Anis, Sahar El-Khoury, and Philippe Bidaud. "An overview of 3D object grasp synthesis algorithms." Robotics and Autonomous Systems 60.3 (2012): 326-336.

Bohg, Jeannette, et al. "Data-driven grasp synthesis-a survey." IEEE Transactions on Robotics 30.2 (2013): 289-309.

Fan, Haoqiang, Hao Su, and Leonidas J. Guibas. "A point set generation network for 3d object reconstruction from a single image." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

Brock, Andrew, et al. "Generative and discriminative voxel modeling with convolutional neural networks." arXiv preprint arXiv:1608.04236 (2016).

Choy, Christopher B., et al. "3d-r2n2: A unified approach for single and multi-view 3d object reconstruction." European conference on computer vision. Springer, Cham, 2016.

Mescheder, Lars, et al. "Occupancy networks: Learning 3d reconstruction in function space." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

Chen, Zhiqin, and Hao Zhang. "Learning implicit fields for generative shape modeling." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

Varley, Jacob, et al. "Shape completion enabled robotic grasping." 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017.

Bajcsy, Ruzena, Yiannis Aloimonos, and John K. Tsotsos. "Revisiting active perception." Autonomous Robots 42.2 (2018): 177-196.

Bohg, Jeannette, et al. "Interactive perception: Leveraging action in perception and perception in action." IEEE Transactions on Robotics 33.6 (2017): 1273-1291.

Kendall, Alex, and Yarin Gal. "What uncertainties do we need in bayesian deep learning for computer vision?." Advances in neural information processing systems. 2017.

Van der Merwe, Mark, et al. "Learning Continuous 3D Reconstructions for Geometrically Aware Grasping." (Under Review) IEEE International Conference on Robotics and Automation (2019).

Wang, Shaoxiong, et al. "3D shape perception from monocular vision, touch, and shape priors." 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018.

Yan, Xinchen, et al. "Learning 6-DoF grasping interaction via deep geometry-aware 3D representations." 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018.