CS 6301 Introduction to Robot Manipulation and Navigation Project Proposal Description

Professor Yu Xiang

September 12, 2022

1 Introduction

For the robotics course project, students can choose a topic related to robotics, and explore the topic in one of the three different ways:

- **Research-oriented.** In this direction, students are going to propose a new idea that has not been explored before in the literature, then implement the new idea and conduct experiments to verify it.
- **Application-oriented.** In this direction, students can apply an existing robotics algorithm or method to a new problem or a new application. For example, if a method is proposed for domain A, the project can explore applying the method to domain B where different data are collected.
- **Implementation-oriented.** In this direction, students can select an existing robotics algorithm or method, and then implement it and conduct experiments to verify the implementation. Since most robotics methods are open-source these days, for implementation-oriented projects, students cannot just use an open-source code and run experiments with it.

For project evaluation, all three categories will be considered equally. A project will be evaluated according its quality in terms of implementation, experiments, presentation and writing, regardless of its category. However, students are encouraged to consider research-oriented projects and application-oriented projects. Even if the introduced novelty is incremental, it is still exploring new things researchers have not been tried before or applying an approach to new applications. Moreover, if you use the Fetch mobile manipulator for experiments, it is possible to try your methods on the real robot in my lab.

2 Proposal Format

The project proposal should be prepared using the the ICRA double column latex format. A useful online LaTex tool is Overleaf `https://www.overleaf.com/`. We have the ICRA latex template accessible here via overleaf: `https://www.overleaf.com/read/rwmhwnwjkrmc`. You can download a copy of the template or make a copy in overleaf for your own project, and then edit it.

The project proposal should be a 1-page PDF using the latex template with the following items:

- **Title**. Let's give a name to your project.
- **Team Members**. List the names of the team members as the authors in the proposal. We expect you to work in groups of 2-4 students for the course projects.
- **Problem Statement**. Describe what is the problem you are trying to solve in this project.
- **Approach**. Describe what is your idea to solve the problem. It is fine if some details have not been figured out in the project proposal. But students should have rough ideas on how to proceed.

  First, explicitly state that which category the project is in: research-oriented, application-oriented or implementation-oriented. Second, for research-oriented projects, describe the proposed idea and the novelty of the idea. For application-oriented projects, describe which approach is going to be used and how to apply this approach to a new application. For implementation-oriented projects, describe which approach is going to be implemented and the plan for the implementation.
- **Simulation Environments and Data**. Describe which simulation environment will be used in the project. If there are existing datasets that can be used for the project, describe what dataset the project is going to use. Students can collect your own datasets from simulation for experiments.
- **Evaluation**. Describe how to evaluate the success of the project. For example, what evaluation metrics will be used to evaluate the performance of the method?
- **References**. Cite related works in the proposal.

3 Suggested Topics

Based on the materials we cover in the lectures, we suggest the follow topics for the course project. However, the scope of the project is not limited to the mentioned topics below. Students can explore other topics in robotics as well. Also, the references in the suggested topics are recent representative works. Students can explore methods beyond these references and propose new ideas for different topics.

- **Model-based Grasping**. Using 3D models of objects, we can first estimate the 6D object pose, i.e., 3D rotation and 3D translation of objects [40, 34, 35, 9]. Then model-based grasp planning [24] and motion planning [5] approaches can be used for robot grasping.
- **Model-free Grasping Planning**. If we cannot get 3D models of objects but we can obtain images or 3D point clouds of objects, grasps can be planned using images or point clouds [2, 21, 6, 25, 32].
- **Reinforcement Learning for Grasping**. Reinforcement learning approaches have been

widely studied for robot grasping. Most approaches focus on using RL for top-down grasping [27, 14, 12, 36, 43]. A few works apply RL to 6D grasping [31, 38, 37].

- **Manipulation of Articulated or Deformable Objects**. In addition to manipulation of rigid objects, active research interests are focusing on manipulation of articulated objects [16, 39, 20, 41] and deformable objects [42, 29, 18, 19].
- **Navigation with ROS Navigation Stack**. You can use the ROS navigation stack `http://wiki.ros.org/navigation` and apply it to robot navigation.
- **Visual Navigation**. Traditional robot navigation focuses on using Lidar to build 2D occupancy maps. Visual navigation studies how to use images as the sensor input for robot navigation. Current approaches focuses on learning navigation policies using images [22, 15, 33, 1].
- **Target-Driven Navigation**. In this problem, a robot is given an image of the destination and the robot needs to navigate to the destination autonomously. A number of learning-based approaches have been proposed for target-driven visual navigation [44, 30, 10, 26].
- **Topological Navigation**. Instead of building 2D or 3D maps for navigation, topological navigation focuses on building topological maps for environments for autonomous navigation [17, 28, 23, 3, 4].
- **Task and Motion Planning (TAMP)**. TAMP focuses on generating task plans and motion plans for robots doing various tasks [13, 8, 11, 7].

4 Robot Simulator Resources

Here are a few robot simulators and simulation environments you can use for the course project:

- Gazebo `https://gazebosim.org/home`. Gazebo is integrated with ROS. If your code runs in Gazebo, it can be easily transferred to a real robot.
- PyBullet `https://pybullet.org/wordpress/`. PyBullet is an easy to use simulator with Python interfaces.
- NVIDIA Isaac Gym `https://developer.nvidia.com/isaac-gym`. Isaac Gym can use GPU acceleration and parallel runs of thousands of environments. It is useful for RL.
- iGibson `https://svl.stanford.edu/igibson/`. iGibson is a simulation environment based on PyBullet. It can be used for robot manipulation and navigation.
- AI2-THOR `https://ai2thor.allenai.org/`, a simulation environment for navigation.
- Habitat `https://aihabitat.org/`, a simulation environment for embodied AI.
- SAPIEN `https://sapien.ucsd.edu/`. SAPIEN contains asserts of articulated objects. It can be useful for articulated object manipulation.
- BulletArm `https://github.com/ColinKohler/BulletArm`, a PyBullet-based simulation environment for benchmarking of several robot manipualtion tasks.

5 Deep Learning Resources

Most recent vision methods leverage deep learning to train neural networks to tackle various problems in robotics. If your project requires training of deep neural networks, you may need to have GPUs for training. **Google Colab** is a great free resources for small amounts of GPU resources: `https://colab.research.google.com/`. Two widely-used deep learning frameworks:

- PyTorch `https://pytorch.org/`
- TensorFlow `https://www.tensorflow.org/`

References

[1] Bo Ai, Wei Gao, David Hsu, et al. Deep visual navigation under partial observability. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 9439–9446. IEEE, 2022.

[2] Jeannette Bohg, Antonio Morales, Tamim Asfour, and Danica Kragic. Data-driven grasp synthesis—a survey. *IEEE Transactions on robotics*, 30(2):289–309, 2013.

[3] Devendra Singh Chaplot, Ruslan Salakhutdinov, Abhinav Gupta, and Saurabh Gupta. Neural topological slam for visual navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12875–12884, 2020.

[4] Kevin Chen, Junshen K Chen, Jo Chuang, Marynel Vázquez, and Silvio Savarese. Topological planning with transformers for vision-and-language navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11276–11286, 2021.

[5] Sachin Chitta, Ioan Sucan, and Steve Cousins. Moveit![ros topics]. *IEEE Robotics & Automation Magazine*, 19(1):18–19, 2012.

[6] Fu-Jen Chu, Ruinian Xu, and Patricio A Vela. Real-world multiobject, multigrasp detection. *IEEE Robotics and Automation Letters*, 3(4):3355–3362, 2018.

[7] Aidan Curtis, Tom Silver, Joshua B Tenenbaum, Tomás Lozano-Pérez, and Leslie Kaelbling. Discovering state and action abstractions for generalized task and motion planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5377–5384, 2022.

[8] Neil T Dantam, Zachary K Kingston, Swarat Chaudhuri, and Lydia E Kavraki. An incremental constraint-based framework for task and motion planning. *The International Journal of Robotics Research*, 37(10):1134–1151, 2018.

[9] Xinke Deng, Yu Xiang, Arsalan Mousavian, Clemens Eppner, Timothy Bretl, and Dieter Fox. Self-supervised 6d object pose estimation for robot manipulation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3665–3671. IEEE, 2020.

[10] Alessandro Devo, Giacomo Mezzetti, Gabriele Costante, Mario L Fravolini, and Paolo Valigi. Towards generalization in target-driven visual navigation by using deep reinforcement learning. *IEEE Transactions on Robotics*, 36(5):1546–1561, 2020.

[11] Caelan Reed Garrett, Rohan Chitnis, Rachel Holladay, Beomjoon Kim, Tom Silver, Leslie Pack Kaelbling, and Tomás Lozano-Pérez. Integrated task and motion planning. *Annual review of control, robotics, and autonomous systems*, 4:265–293, 2021.

[12] Stephen James, Paul Wohlhart, Mrinal Kalakrishnan, Dmitry Kalashnikov, Alex Irpan, Julian Ibarz, Sergey Levine, Raia Hadsell, and Konstantinos Bousmalis. Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12627–12637, 2019.

[13] Leslie Pack Kaelbling and Tomás Lozano-Pérez. Integrated task and motion planning in belief space. *The International Journal of Robotics Research*, 32(9-10):1194–1227, 2013.

[14] Dmitry Kalashnikov, Alex Irpan, Peter Pastor, Julian Ibarz, Alexander Herzog, Eric Jang, Deirdre Quillen, Ethan Holly, Mrinal Kalakrishnan, Vincent Vanhoucke, et al. Scalable deep reinforcement learning for vision-based robotic manipulation. In *Conference on Robot Learning*, pages 651–673. PMLR, 2018.

[15] Peter Karkus, Shaojun Cai, and David Hsu. Differentiable slam-net: Learning particle slam for visual navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2815–2825, 2021.

[16] Dov Katz and Oliver Brock. Manipulating articulated objects with interactive perception. In *2008 IEEE International Conference on Robotics and Automation*, pages 272–277. IEEE, 2008.

[17] Kurt Konolige, Eitan Marder-Eppstein, and Bhaskara Marthi. Navigation in hybrid metric-topological maps. In *2011 IEEE International Conference on Robotics and Automation*, pages 3041–3047. IEEE, 2011.

[18] Rita Laezza, Robert Gieselmann, Florian T Pokorny, and Yiannis Karayiannidis. Reform: A robot learning sandbox for deformable linear object manipulation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4717–4723. IEEE, 2021.

[19] Xingyu Lin, Zhiao Huang, Yunzhu Li, Joshua B Tenenbaum, David Held, and Chuang Gan. Diffskill: Skill abstraction from differentiable physics for deformable object manipulations with tools. *arXiv preprint arXiv:2203.17275*, 2022.

[20] Liu Liu, Wenqiang Xu, Haoyuan Fu, Sucheng Qian, Qiaojun Yu, Yang Han, and Cewu Lu. Akb-48: A real-world articulated object knowledge base. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14809–14818, 2022.

[21] Jeffrey Mahler, Jacky Liang, Sherdil Niyaz, Michael Laskey, Richard Doan, Xinyu Liu, Juan Aparicio Ojea, and Ken Goldberg. Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. *arXiv preprint arXiv:1703.09312*, 2017.

[22] Bar Mayo, Tamir Hazan, and Ayellet Tal. Visual navigation with spatial attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16898–16907, 2021.

[23] Xiangyun Meng, Nathan Ratliff, Yu Xiang, and Dieter Fox. Scaling local control to large-scale topological navigation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 672–678. IEEE, 2020.

[24] Andrew T Miller and Peter K Allen. Graspit! a versatile simulator for robotic grasping. *IEEE Robotics & Automation Magazine*, 11(4):110–122, 2004.

[25] Arsalan Mousavian, Clemens Eppner, and Dieter Fox. 6-dof graspnet: Variational grasp generation for object manipulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2901–2910, 2019.

[26] Yiding Qiu, Anwesan Pal, and Henrik I Christensen. Target driven visual navigation exploiting object relationships. *arXiv preprint arXiv:2003.06749*, 2(7), 2020.

[27] Deirdre Quillen, Eric Jang, Ofir Nachum, Chelsea Finn, Julian Ibarz, and Sergey Levine. Deep reinforcement learning for vision-based robotic grasping: A simulated comparative evaluation of off-policy methods. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6284–6291. IEEE, 2018.

[28] Nikolay Savinov, Alexey Dosovitskiy, and Vladlen Koltun. Semi-parametric topological memory for navigation. *arXiv preprint arXiv:1803.00653*, 2018.

[29] Daniel Seita, Pete Florence, Jonathan Tompson, Erwin Coumans, Vikas Sindhwani, Ken Goldberg, and Andy Zeng. Learning to rearrange deformable cables, fabrics, and bags with goal-conditioned transporter networks. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4568–4575. IEEE, 2021.

[30] Shamane Siriwardhana, Rivindu Weerasekera, and Suranga Nanayakkara. Target driven visual navigation with hybrid asynchronous universal successor representations. *arXiv preprint arXiv:1811.11312*, 2018.

[31] Shuran Song, Andy Zeng, Johnny Lee, and Thomas Funkhouser. Grasping in the wild: Learning 6dof closed-loop grasping from low-cost demonstrations. *IEEE Robotics and Automation Letters*, 5(3):4978–4985, 2020.

[32] Martin Sundermeyer, Arsalan Mousavian, Rudolph Triebel, and Dieter Fox. Contact-graspnet: Efficient 6-dof grasp generation in cluttered scenes. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13438–13444. IEEE, 2021.

[33] Varun Tolani, Somil Bansal, Aleksandra Faust, and Claire Tomlin. Visual navigation among humans with optimal control as a supervisor. *IEEE Robotics and Automation Letters*, 6(2):2288–2295, 2021.

[34] Jonathan Tremblay, Thang To, Balakumar Sundaralingam, Yu Xiang, Dieter Fox, and Stan Birchfield. Deep object pose estimation for semantic robotic grasping of household objects. *arXiv preprint arXiv:1809.10790*, 2018.

[35] Chen Wang, Danfei Xu, Yuke Zhu, Roberto Martín-Martín, Cewu Lu, Li Fei-Fei, and Silvio Savarese. Densefusion: 6d object pose estimation by iterative dense fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3343–3352, 2019.

[36] Dian Wang, Robin Walters, Xupeng Zhu, and Robert Platt. Equivariant q learning in spatial action spaces. In *Conference on Robot Learning*, pages 1713–1723. PMLR, 2022.

[37] Lirui Wang, Xiangyun Meng, Yu Xiang, and Dieter Fox. Hierarchical policies for cluttered-scene grasping with latent plans. *IEEE Robotics and Automation Letters*, 7(2):2883–2890, 2022.

[38] Lirui Wang, Yu Xiang, Wei Yang, Arsalan Mousavian, and Dieter Fox. Goal-auxiliary actor-critic for 6d robotic grasping with point clouds. In *Conference on Robot Learning*, pages 70–80. PMLR, 2022.

[39] Ruihai Wu, Yan Zhao, Kaichun Mo, Zizheng Guo, Yian Wang, Tianhao Wu, Qingnan Fan, Xuelin Chen, Leonidas Guibas, and Hao Dong. Vat-mart: Learning visual action trajectory proposals for manipulating 3d articulated objects. *arXiv preprint arXiv:2106.14440*, 2021.

[40] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. *arXiv preprint arXiv:1711.00199*, 2017.

[41] Zhenjia Xu, Zhanpeng He, and Shuran Song. Universal manipulation policy network for articulated objects. *IEEE Robotics and Automation Letters*, 7(2):2447–2454, 2022.

[42] Jihong Zhu, Andrea Cherubini, Claire Dune, David Navarro-Alarcon, Farshid Alambeigi, Dmitry Berenson, Fanny Ficuciello, Kensuke Harada, Xiang Li, Jia Pan, et al. Challenges and outlook in robotic manipulation of deformable objects. *arXiv preprint arXiv:2105.01767*, 2021.

[43] Xupeng Zhu, Dian Wang, Ondrej Biza, Guanang Su, Robin Walters, and Robert Platt. Sample efficient grasp learning using equivariant models. *arXiv preprint arXiv:2202.09468*, 2022.

[44] Yuke Zhu, Roozbeh Mottaghi, Eric Kolve, Joseph J Lim, Abhinav Gupta, Li Fei-Fei, and Ali Farhadi. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 3357–3364. IEEE, 2017.