Shape Reconstruction and Tactile Exploration with Machine Learning Paradigms

Burak Ömür   Hewei Gao

Technical University of Munich

I. Objective

In the last years, the need for 3D shape understanding with the expansion in robotics has vastly increased. Object geometry is crucial when aiming for grasping and manipulation, [9] and many existing robots are already designed to act on objects. Previous works tackled this shape understanding problem using single-view and multi-view images (RGB), depth-mapped images (RGB-D), and tactile readings. The current scenario is moving this approach further and integrating the active-sensing approach [8] which is the tactile exploration task. The task of tactile exploration is developing the best strategy to infer an object's complete shape. Therefore, one needs a strategy to gather tactile points and this strategy must be reasonable to comprehend the shape of the object. This task aimed to strengthen entities that rely on their vision and their exploration skills. For example, a robot arm learning to grasp objects with various shapes could make use of the task.

We plan to start on 2D to uncover best practices and then move the idea to 3D. Our objective is, first, the construction of a 2D dataset by utilizing a 3D one, (i.e. ShapeNet [2]). Second, training of a neural net on this 2D dataset for shape completion. Then, if it generalizes well on the target domain; third, utilization of this model as a reward function for a reinforcement learning (RL) agent to obtain an exploration strategy. Only some of the previous works tackled this tactile exploration task using RL [4], instead they used Gaussian Processes and Kalman Filters [3], [5], [6], and therefore this learning paradigm is yet to be discovered.

II. Related Work

The idea of combining sensory information from vision and tactile sensors for the reconstruction task is not new. Ilonen et al. [5] proposed an optimal estimation method for combining visual and tactile data in order to reconstruct a complete 3-D model of an unknown object when the object is grasped. However, the robot in this work executed poke and grasp only once, which means they didn't discuss how to determine the best next tactile exploration based on previous points. The work of Watkins-Valls et al. had a similar situation. [9] They provided an architecture that incorporates depth and sparse tactile information to reconstruct 3D models for robot grasp tasks and outperformed other visual-tactile approaches. While this work applied CNN fed with both depth and tactile information, which is consistent with our objective, they randomly sampled 40 points to generate tactile data rather than using a learning exploration strategy. This problem received attention in the work of J.Smith et al. [8], which proposed a data-driven framework to guide shape exploration. They set four experimental scenarios and compared the results of five different exploration policies. However, in the final experiment result, the policy using the RL method failed to achieve the best grade in any experimental scenario.

III. Technical Outline

Dataset construction and visualization: There are many existing 3D model datasets in the literature, ShapeNet [2], REGRAD [10], YCB [1]. Our aim is to create our own 2D shape dataset using ShapeNet. In this mission, we want to project 3D objects onto the 2D domain from many different viewpoints. In this course, we will use the signed-distance-field (SDF) approach to create our own grids. Then, we will visualize our dataset and make our it available in our repository. [1]

2D-shape understanding model: In this step we need a model to infer a 2D object's shape from a portion of the image and a couple of sparse tactile points sampled with some deterministic approach. This shape prediction task is similar to the segmentation task where the aim is to predict a class for each pixel in the image. Therefore, we prefer to use a U-Net [7] architecture to output shape predictions with high resolution.

Training: In training, we will utilize our projected 2D dataset on our built model. We will first test our architecture integrity via overfitting on a small portion of the dataset, then if it passes our test, we will use our complete dataset to train our 2D shape prediction model.

Obtaining tactile strategies: After obtaining a successful shape prediction model, we want to further utilize the learning paradigm and use reinforcement learning to output tactile exploration strategies. We plan to use our trained model as a reward function on our explorer agent. Our aim is to formalize our agent similar to the Recurrent Attention Model like in [4]. Then, training our agent using again our 2D dataset and tactile strategy obtained from it. In the end, we will visualize and present the results.

Optional, bring shape understanding model to 3D fusing visual and tactile together: Optionally, we plan to switch to 3D if we have enough incentives originating from the results of the 2D experiment. In 3D, the aim and procedures will be the same except for the dataset construction part which we already have in 3D.

---

[1]https://github.com/darktheorys/ADL4R-tactile-exploration

REFERENCES

[1] Berk Calli, Aaron Walsman, Arjun Singh, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M. Dollar. Benchmarking in manipulation research: Using the yale-CMU-berkeley object and model set. *IEEE Robotics &amp Automation Magazine*, 22(3):36–52, sep 2015.

[2] 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.

[3] Danny Driess, Daniel Hennes, and Marc Toussaint. Active multi-contact continuous tactile exploration with gaussian process differential entropy. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 7844–7850, 2019.

[4] Sascha Fleer, Alexandra Moringen, Roberta L. Klatzky, and Helge Ritter. Learning efficient haptic shape exploration with a rigid tactile sensor array. *PLOS ONE*, 15(1):e0226880, January 2020.

[5] Jarmo Ilonen, Jeannette Bohg, and Ville Kyrki. Fusing visual and tactile sensing for 3-d object reconstruction while grasping. In *2013 IEEE International Conference on Robotics and Automation*, pages 3547–3554, 2013.

[6] Nawid Jamali, Carlo Ciliberto, Lorenzo Rosasco, and Lorenzo Natale. Active perception: Building objects' models using tactile exploration. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pages 179–185, 2016.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[8] Edward J. Smith, David Meger, Luis Pineda, Roberto Calandra, Jitendra Malik, Adriana Romero, and Michal Drozdzal. Active 3d shape reconstruction from vision and touch. 2021.

[9] David Watkins-Valls, Jacob Varley, and Peter Allen. Multi-modal geometric learning for grasping and manipulation, 2018.

[10] Hanbo Zhang, Deyu Yang, Han Wang, Binglei Zhao, Xuguang Lan, Jishiyu Ding, and Nanning Zheng. Regrad: A large-scale relational grasp dataset for safe and object-specific robotic grasping in clutter, 2021.