A Contact Exploitative Approach to the Amazon Robotics Challenge

Eadom Dessalene[1], Georgios Georgakis[1], Md. Reza[1], Yimeng Li[1], Yossi Ovcharik[3], Amir Shapiro[3], Jana Košecká[1], Daniel Lofaro[2]

***Abstract*— The goal of the Amazon Robotics Challenge is to automate pick and place operations in unstructured environments by applying the state of the art in robotics. The two areas of complexity that were expanded relative to previous competitions are in storage density and the presence of unknown objects prior to the competition. To deal with these two factors, we use two contact-exploitative modes of manipulation along with the use of a soft GelSight [1]-inspired elastomer used to prevent possible toppling of objects. We present an overview of our approach to the picking task of the Amazon Robotics Challenge as follows: We begin with our choice of hardware, and then discuss our deep learning based vision approach. We then go into our modes of operation along with the situations they are used in and conclude with the limitations of our chosen approach.**

I. Hardware

Our choice of robot is the Baxter Research Robot along with a Dataspeed Inc. mobile base. Two RGB-D cameras are attached to each wrist, enabling a multitude of opposing viewpoints for each bin. Unlike the vast majority of most other teams, our choice of end effector is the Allegro Hand, a 20 DOF Allegro Hand capable of grasping every single object available in the known dataset.

A. Storage System

Our customized storage system is unique in that every horizontal surface of the shelf is coated with a soft, GelSight-inspired elastomer capable of transmitting tactile images. Our elastomer of choice is a simple 3:2 ratio of TPE (Thermoplastic Elastomer) to Toluene, along with a thin coating of gray pigment particles on the surface. Tactile images, along with their RGB and Depth correspondences are shown in Figure 1.

While the elastomer can be made softer by increasing the volume of toluene, we choose for a stiffer Shore 00-30 hardness value such that while the resting state of the objects does not transmit any clear images (the dumbell being the exception), the robot exerting a force on an object against the elastomer transmits a clear tactile image. This is to avoid the object instability that comes with a near-viscous layer.

II. Object Detection

We follow the latest trends in object detection and employ a Deep Convolutional Neural Network (DCNN) for localizing and recognizing the objects in the bins and tote. Recently, models such as Faster R-CNN [5] produced state-of-the-art results on popular benchmarks and last year's challenge [8]. This model employed an end-to-end training where both the localization and the classification of the objects is considered by the loss function. One of the shortcomings of these approaches is the fact that they require large amount of bounding box annotations for training. For the competition, only a few cropped images of each object are available which is not sufficient to train the entire model end-to-end. We plan to address this issue by first training a model with a large synthetic dataset of objects superimposed in real background scenes at informatively chosen positions and scales, and then fine-tuning the model with a small number of manually annotated images of the competition objects inside the bins and tote. For the generation of the synthetic dataset, we plan to use 2D images of object models from BigBird [2], previous Amazon competitions [4], and background scenes from the NYU-V2 dataset [3]. Details on how the synthetic dataset is generated can be found in [6]. Example images from the dataset are shown in Figure 3.

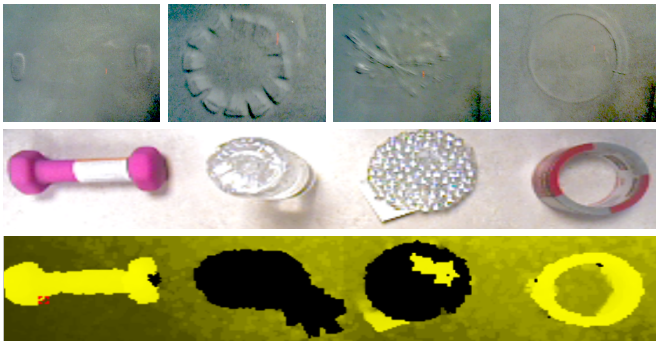

Fig. 1: Tactile imaging for four of the known ARC objects, along with their rgb and depth equivalents. As seen above, tactile imaging is fully capable of transmitting detail unavailable to traditional time-of-flight depth sensors.

III. Manipulation

After localizing each of the target objects in the chosen bin, the swept volume of each object over the depth of the bin is collected. This swept volume represents a linear retraction of each object over the depth of the bin. Collision checks are counted between each of the swept volumes, and the target object that prevents the highest number of linear retractions of the surrounding objects is chosen for grasping. Naturally, it is objects near the brim of the bin that will first be chosen for grasping as they are most likely to offend

[1] Computer Science Department, George Mason University 4400 University Drive, VA, USA

[2]Electrical and Computer Engineering Department, George Mason University 4400 University Drive, VA, USA

[3]Mechanical Engineering Department, Ben-Gurion University of the Negev, Be'er Sheva, 8499000, Israel

Fig. 2: The slide to edge manipulation process is illustrated above. After the object is dragged past the brim of the shelf, the hand is retracted and a pre-selected hand template is moved a set distance away from the brim of the shelf such that a compliant grasp will successfully grasp any object slightly portruding from the brim.

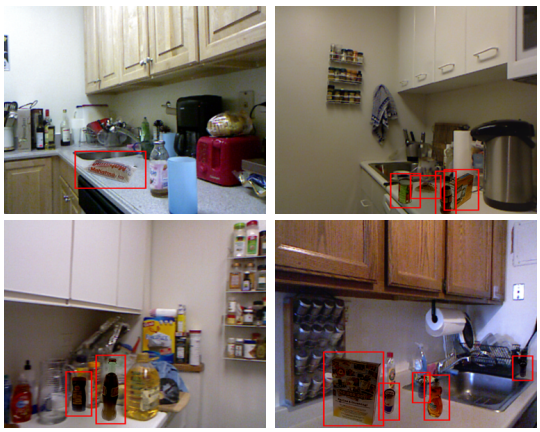

Fig. 3: Example images from the generated synthetic dataset. The superimposed objects are marked with red bounding boxes.

the retraction of objects behind them. Rather than resort to a static grasp planner, we draw from the work of [10] and utilize *Surface Constrained Grasping* and *Slide to Edge Grasping*. The mode of operation to choose is determined by fitting a plane bounded by the size of the 3 fingers of the Allegro Hand to the surface of the mesh of the target object using RANSAC [9].

When a plane is found, a line is drawn perpendicular to the center of the plane: should the line be longer than a threshold that prevents the thumb of the hand from grasping the opposing side, the plane is rejected. In our case, the maximum allowable threshold for the Allegro Hand is 5 cm. Additionally, this threshold filters the sliding of any physically unstable objects as the objects chosen for sliding are likely to be extremely flat. Otherwise, the search continues. When a plane is found, a *Slide to Edge Grasp* is executed. Otherwise, a *Surface Constrained Grasp* against the support surface the target object is lying on is executed.

A. Reach

The reaching step is comprised of two steps: The execution of a Constrained BiDirectional RRT (CBiRRT) [7] when the hand is approaching the relevant bin and a local Jacobi controller when the hand is operating from within the bin. The search space of the CBiRRT is the configuration space of the arm before contact is made. When exerting a force on any object for initializing a *Slide to Edge Grasp*, the Jacobi controller is executed in force control against the surface of the wall.

B. Surface Constrained Grasp

The compliant finger-tip placement grasp is executed with a compliant grasp against the surface the target object is lying against. Initially, the support surface the object is lying against is identified. Unoccluded regions of the support surface are segmented, and the placement of the thumb is sampled in this unoccluded region surrounding the target object. The placement of the thumb is executed using a stored hand template, and a compliant grasp with the opposing three fingers is executed as their fingertip trajectories traverse the support surface until the object is fully enclosed within the aperture of the hand.

C. Slide to edge Grasping

Slide to edge grasping is a form of hand closing after the robot has moved the object over the brim of the shelf, effectively exposing an additional side of the object for grasping. Figure 2 illustrates this process. After identifying where to put the fingers for sliding, a *reach* is executed. Using the local Jacobi controller, a force is executed perpendicular to the support surface such that the center of friction of the object is pulled towards the wall, enabling easier sliding. A slow linear retraction of the object against the wall using the Jacobi controller is performed, until the Allegro Hand is completely out of the shelf, or until the object begins to topple as detected by the tactile feedback from the shelf, explained in IV.

However, only the horizontal support surfaces of the shelf are coated with the elastomer, meaning tactile images are only available when a downward force is exerted against the gel. While tactile images are always available when sliding against the horizontal support surfaces, slide-to-edge manipulation will require a slight downward force vector when sliding across vertical support surfaces such that the elastomer can track the tactile images of the target object.

D. Extract Out of Bin

After grasping the target object with a surface constrained grasp, the simulation of a slow linear retraction is simulated by the combined swept volume of the hand and object over the depth of the bin. When iterating over 6D pose re-orientations of the hand, each iteration is checked for a swept volume that avoids collision with the shelf and

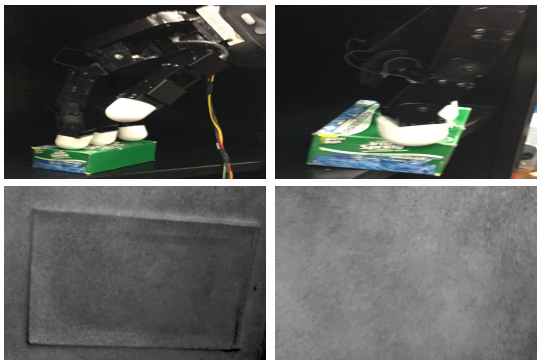

Fig. 4: Tactile images collected from a prototype elastomer. It is shown that as the object is pressed against a planar surface, a clear tactile image is transmitted. During peaks of instability as the object begins to tip, the tactile image reverts to its default.

the surrounding clutter. When a collision-free swept volume is found, the arm checks for kinematic feasibility of the pose re-orientation. If successful, the arm joint commands are executed and the mobile base moves in a -z direction, extracting the object out of the bin in a collision free manner.

IV. Feedback

The slide to edge manipulation step performed open-loop can prove very dangerous, as a slight tip between the object and the brim of the shelf can lead to the robot mistakenly throwing the object outside the storage system, with no direct means of retrieval. The presence of a tactile tracking system successfully closes this loop. The transition from a full projection of the face during the sliding process to an empty tactile image during unstable transitions is shown in Figure 4. Simple edge detection suffices to determine the stability of the dragged object. It was because of the need for a high FPS tactile tracking system that we opted for responsive 2D imaging as opposed to a drastically reduced FPS rate that comes with using photometric stereo for 3D imaging.

V. Limitations

While our approach to the challenge is robust to clutter and occlusions, there are several rules to how the objects must be stowed inside the bins.

- Planar objects to be slid must be near parallel to the support surface they are lying against.
- Planar objects to be slid must be present at the front of each surface of the bin, and all other objects must be on top of these planar objects or at the back of the shelf.
- Careful placement of the objects must be performed such that the robot can grasp the objects and without necessary whole-arm movement, extract the object out of the bin with the movement of the mobile base.

VI. Conclusion

In this paper we have discussed the use of a tactile sensing shelf along with the use of two contact-rich modes of manipulation. Our approach to the ARC has multiple future avenues of exploration. One future direction we are excited towards is the use of optical flow tracking from the tactile imaging to detect possible slippage between the hand and the object. We look forward to participating in the 2017 ARC event.

References

[1] Li, Rui, and Edward H. Adelson. "Sensing and recognizing surface textures using a gelsight sensor." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013.

[2] A. Singh, J. Sha, K. Narayan, T. Achim, and P. Abbeel, *A large-scale 3D database of object instances*, in IEEE International Conference on Robotics and Automation (ICRA), 2014.

[3] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, *Indoor segmentation and support inference from RGB-D images*, ECCV, 2012.

[4] http://www.robocup2016.org/en/events/amazon-picking-challenge/

[5] S. Ren, K. He, R. Girshick, and J. Sun, *Faster R-CNN: Towards real-time object detection with region proposal networks*, NIPS, 2015.

[6] G. Georgakis, A. Mousavian, A. C. Berg, and J. Kosecka, *Synthesizing Training Data for Object Detection in Indoor Scenes*, arXiv:1702.07836, 2017.

[7] Berenson, Dmitry, et al. "Manipulation planning on constraint manifolds." Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, 2009.

[8] C. Hernandez, M. Bharatheesha, W. Ko, H. Gaiser, J. Tan, K. Deurzen, M. Vries, B. Mil, J. Egmond, R. Burger, M. Morariu, J. Ju, X. Gerrmann, R. Ensing, J. Frankenhuyzen, and M. Wisse, *Team Delft's Robot Winner of the Amazon Picking Challenge 2016*, arXiv:1610.05514, 2016.

[9] Tarsha-Kurdi, Fayez, Tania Landes, and Pierre Grussenmeyer. *Hough-transform and extended ransac algorithms for automatic detection of 3d building roof planes from lidar data.* Proceedings of the ISPRS Workshop on Laser Scanning. Vol. 36. 2007.

[10] Eppner, Clemens, et al. *Exploitation of environmental constraints in human and robotic grasping.* The International Journal of Robotics Research (2015): 0278364914559753.