ML-Assisted Fruit-Sorting Bot

Learning fruit manipulation using Machine Learning Methods

Kristian Praizner
Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA
krisp@mit.edu

John Moler
Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA
johnm2@mit.edu

***Abstract*—The agricultural industry is the backbone of society. This project proposes a machine learning-assisted robotic system using the KUKA IIWA to sort various fruits into designated bins, benefiting the industry. The system addresses three key challenges: perception, motion planning, and grasping.**

Perception employs machine learning algorithms and point cloud processing to classify fruits by type, shape, and size. Motion planning utilizes Drake's trajectory optimization for precise kinematic control. Grasping incorporates antipodal heuristics and models fruit malleability to handle fruits gently, minimizing damage.

Inspired by prior work on robotic grippers and CNN-based object pose estimation, the project integrates cutting-edge robotics and machine learning techniques to deliver a scalable, efficient fruit-sorting solution for real-world agricultural applications.

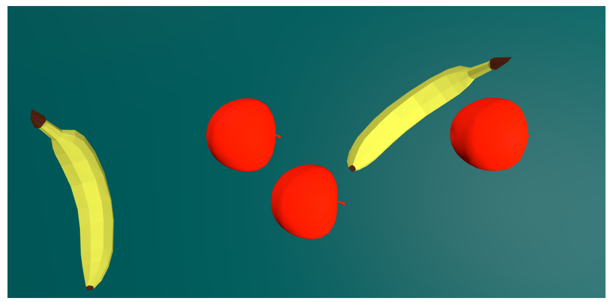

Fig. 1. Example of a cluttered fruit scene.

I. Introduction

Sorting fruits manually is labor-intensive, time-consuming, and prone to errors [1]. Automation can increase the speed and accuracy of sorting, significantly reducing labor costs and the potential for human error, which is vastly integral in industries with tight deadlines such as the agricultural industry. Once implemented, a robotic system can be scaled easily to handle varying workloads without increasing human resources. Reducing human interaction with heavy lifting or repetitive tasks such as sorting will lead to fewer workplace injuries [2], higher efficiency (more fruit sorted per unit time), and potentially higher accuracy. The robotic system can be optimized to pick certain fruits based off a vision model, making it able to spot deformities and act more diligently than a trained human. The system can also be trained to optimize its gripping force, protecting the deformable fruits from damage.

II. Related Works

A. Sorting Clutter

Manipulating objects in a crowded scene is a common problem in robotic manipulation tasks. This task requires motion planning, grasping, and point cloud processing. The in-class solution to this problem is riddled with bugs and is not robust at all. The motion planning system does not try to avoid obstacles, and the vision system doesn't identify objects but only looks at point clouds and tries to pick them. We use this base model as a starting point, and build off of it to make it considerably more robust and efficient. [4] Goes into a similar idea of using pose estimation with a CNN-based approach that retrains itself as it categorizes objects. While this novel approach is quite effective for their solution, it proves to be too resource-heavy and complex to quickly manipulate fruit in a factory setting. Therefore, a simpler solution for training a network based on the fact that we know what fruits are going to be in the scene is more effective.

B. Deformability

Manipulation of deformable objects is a challenging task in robotics. [3] discusses many important subjects related to our project. First, it discusses possible simulation methods for simulating fruit damage and bruising, with pressure feedback being one of the main drivers. We are advancing on that by using machine feedback and object deformity as measures so that we can also pick up fruits like bananas that might get damaged but not bruise as easily. It also talks about the risks of different types of damage with different gripper types, where a basic electric gripper can be used fairly effectively with low damage risk.

III. METHODOLOGY

We developed a KUKA IIWA robotic system capable of sorting different types of fruits into their respective bins. This project's details can be divided into three different subparts, including the setup, the technical aspects of the robot, and the evaluation criteria. The technical aspect can also be divided into three subproblems: perception, motion planning, and grasping.

A. Setup

The scene is modeled to represent a realistic factory setting. A pile of assorted fruits is cluttered on a table to represent a conveyor belt. The IIWA arm sits on the side of the table and has bins on the other side of it that correspond to the different types of fruit being modeled. These fruits are deformable as a way to simulate damage to them during the process. The fruit will be of different types, shapes, and sizes. There also exist cameras in the scene that help with capturing the point clouds through three different angles

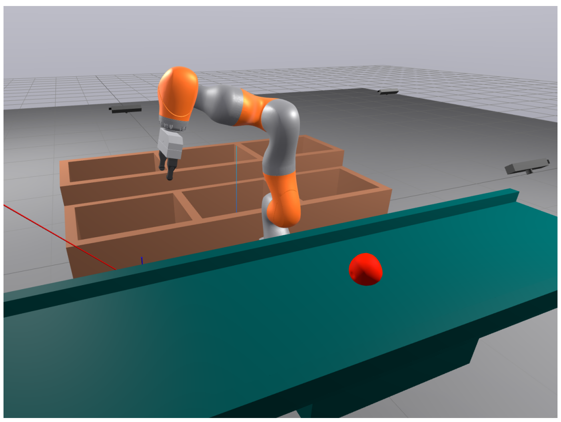

Fig. 2. Initial scene showing the robot with bins behind it. Also with the red apple on the track on front of it and cameras above the track.

B. Technical

- **Perception**
  The crux of the problem comes from the robot's requirement to identify the different fruits that appear before it. In order to achieve this, we used data augmentation in order to generate bulk data sets to train an RCNN algorithm. In order to generate a random dataset for training, we used data augmentation to create many different random fruit configuration and extract mask images, rgb images, and labels to train on. After generating many random configurations of fruits, we then fed them into an RCNN algorithm, which is trained to then identify objects and provide predicted masks given an RGB image. We then place this trained algorithm back into the simulation environment to assist in our sorting task. This helps the grasping module because it allows the isolation and segmentation of the item masks, and therefore the isolation of their point clouds to assist in grasp selection. We also get a label from this algorithm, which the motion planning module uses in order to select which bin the fruit should be placed in. A key aspect of this system is that it is capable of function in a normal application where the data the robot gets purely RGBD. With only a picture and a depth map, the robot can find the mask of the fruit, and isolate the point clouds based on that information.
  Side Note: Though it is not implemented yet because of technical difficulties implementing grasp control, the plan was to use the random configuration creator from the perception module to create scenarios for testing the full end to end implementation.
- **Motion Planning**
  Once the fruit has been identified, our robot's motion control system guides the robotic arm and end effector to pick up the fruit and place it in its corresponding bin. This requires precise kinematic control, and optimal path planning from Drake's trajectory optimization functionality and rrt planning to ensure efficient and safe movement. This motion module can be split into separate stages for our different movement stages. They are as follows: prepare to pick up fruit (position hand in close proximity to expected location), pick up the fruit (lower end effector to fruit level), and move fruit into their respective baskets. Also incorporated in this Motion Planning is collision avoidance, as we do not want the arm to hit any other objects while in motion.
- **Grasping**
  Starting from the end effector position achieved by our motion planner and the point clouds from the cameras in the perception module, the built-in IIWA arm grasper selects where we should make contact with the object to enable a firm grasp. We used antipodal grasping heuristics in the perception module to enable a valid grasp. This module also accounts for the malleability and fragility of fruit, and is trained to grasp in a gentle way so as to not damage or alter the fruit. Without this, this project would lack real world applications because our robot would end up damaging the fruits.

- **Deformability**
  Drake's built in support for deformability is used in order to simulate damage done to fruits by the robot arm. This module allows us to assess how viable the arm is in actual real-world applications, as the highest priority in applications like this will be to preserve the produce until it reaches the end consumer. Once the fruit models are acquired, they will be calibrated to make them roughly as rigid as their bodies would be in real-world applications. We have provided a function in our code that initialize a vtk file into our simulation, allowing new fruit to be added into the variations very easily. During the simulation, the way we are going to observe deformability is through the point clouds of the fruit. By observing the difference in the point clouds before grasping and motion and after grasping and planning, we can predict how much these fruits would be damaged should we employ this system in a real-world environment. A limitation of note is that these files need to be calibrated manually in order to get accurate results. For our simulation, we used real world mass metrics, and adjust physics coefficients based

on real world experiments. This is a complex and long-winded process of trial and error so this is a spot where our project could be expanded to be more accessible.

C. Evaluation

Evaluation of the robotic system is dependent on various criteria. These criteria are speed (fruit sorted per unit time), accuracy (percentage of fruit correctly placed in bins), and damage to the fruit (measured by deformations). Our main goal is to maximize accuracy, with fruit damage and speed being secondary values we would like to maximize. Our initial goal is 90 percent accuracy, and that value may increase or decrease depending on how our first trial runs go. Goals for the damage and speed are determined after an initial project run has been completed.

IV. EVALUATION AND DISCUSSION

A. Perception

Due to technical limitations in a deep-note environment, datasets were only able to generate 100 data points at a time. This likely reduced the effectiveness of our perception module, but we are proud to say that we still achieved a very successful model with our limited data, potentially due to our smart data generation.

By using a pre-trained item identification Mask-RCNN model, we were able to extract a lot of use out of our data we were able to generate. When applied to random configurations of fruit (excluding edge cases where fruit falls out of frame), the algorithm we trained was able to achieve a 93 percent accuracy when observing a fruit configuration from a single angle. Since we are using three cameras from different angles in the actual grasp selection, we expect the results to be even more exceptional should our motion planning and grasp module achieve full functionality.

We can use this metric as a proxy for our accuracy metric, and if we do, we get a very good result for our initial project. As a reminder, we are limited in the number of data points and the number of training epochs we are able to perform, so 93 percent is a very good starting point. Given proper training resources and a number of data points closer to 10,000 or 100,000, we should be able to achieve close to 99 percent accuracy in object identification and mask segmentation and therefore a close to 99 percent accuracy in bin sorting given a good motion planning module.

To summarize, we blow our accuracy goal out of the water, which is nice to see. In edge cases, our accuracy can drop below 88 percent, but most of the time, we are well above 90 percent.

Author's Note: The model was trained on a CPU so the training was slow and there was not much of it. A technical accident resulted in the original machine that our team was using catching fire, and therefore we operated on limited capacity for the final stages of our project. Our team judges our current accuracy to be exceptional and predict our full functionality would be accurate enough for use in real-world applications.

B. Motion Planning

The RRT implementation of the motion planning successfully brought the end-effector to the necessary points without collision, however the speed at which it worked was not up to the initial standards. Further optimization methods are needed to increase this speed to make it suitable for a factory environment.

C. Grasping

The grasping module - inherently to the structure of the system - is dependent on the perception module. Due to unforeseen technical difficulties, the fully combined grasping module was not working properly by the time that this was written. Due to the concerns in the other issues, a simpler grasping metric with point cloud processing with ICP was done for proof of concept, and valid grasps were achieved on the simple models. Although the entire full-stack system could not come together to produce valid force-controlled grasps on deformable objects, simple grasps were achieved for the sake of the paper.

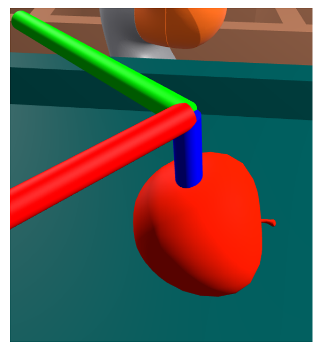

Fig. 3. Example of a valid grasp pose on a apple. Pose is represented by a Meshcat triad.

D. Deformability

Unfortunately, due to technical difficulties related to the vtk file format and the difficulty of creating or converting properly formatted files, we were unable to find or create proper vtk files for our fruit. Even after receiving support from TAs, our simulation environment simply could not process the fruit vtk files that we were composing. However, we were able to run the simulation using the drake torus vtk file. We used this to provide a proof-of-concept for our idea, even if we were unable to get our actual fruit models working. By editing the physics coefficients, we were able to provide different behaviors to the models to simulate different densities and material behavior. Although we were unable to get our grasping and motion planning working to really see to what extent our robot can sort fruit without extra damage, we were able to see what it would look like. (Fig 4)

V. FUTURE WORK

A. Perception

It would be interesting to expand this project in two ways: we want more training data in more varied environments.

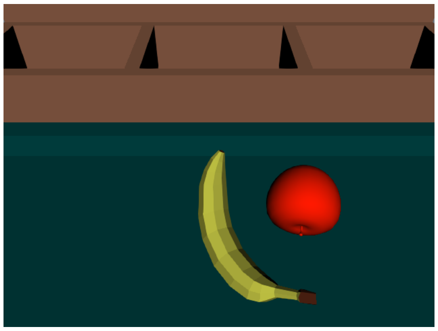

Fig. 4. Example of a fruit configuration used for training.

If we were to expand this with more training data, we would mainly be wanting higher accuracy (which occurs naturally with more data) and more fruit variations. Currently we have two fruit that are not very similar (apples and bananas). An interesting expansion would be to add fruit variations that look similar, such as peaches and apples in the same simulation group. This would be closer to real world applications where perhaps these machines will be used in apple orchards and may need to separate apples, pears, and peaches.

Another direction would be to train this in more varied environments. Our current model is trained specifically on our "conveyor belt" environment. This environment is constant in our model and we do not have any other environment in which to train. This will almost definitely make our model less reliable if used in different spaces. It would be interesting to see this model trained in varied environments to make it more easily integrated into other systems.

B. Deformability

Due to unforeseen limitations with the codebase given for vtk conversions, we were unable to provide performance metrics for our deformability module. However, our proof-of-concept provides room for expansion on future projects. A nice short expansion would be to make the process of new models more automated so that they do not have to be manually calibrated. Our function for integration of deformable bodies enables future projects to alter the parameters for their own purposes, which enables the potential to abstract this project to many different applications. If the goal is to harvest the produce at a different ripeness level, this aspect will make it easy for the model to be adjusted. Upon adjustment, the model would be more gentle if the fruit is supposed to be softer or more ripe.

C. Grasping

The current grasp solution does not work with the full-scale system, so future work can be fully completing this module with additional costs involved with the fruits' malleability. Once a full grasp solution is found, additional vision elements can be expanded on to make this fully work in a moving conveyor belt, as was envisioned in the original proposal of the project. Along with the speed constraints with the motion planning, this would make the system efficient enough to be employed in a realistic setting.

VI. CONCLUSION

This project has presented a machine learning-assisted robotic fruit-sorting system designed to address three key challenges in automated agricultural tasks: perception, motion planning, and gentle grasping. By combining a KUKA IIWA robotic arm with advanced vision techniques, deformability simulations, and intelligent manipulation strategies, we have taken important steps toward a more robust and flexible solution suitable for real-world industrial applications. The system, once fully fleshed out, could potentially operate in a factory setting, increasing efficiency, reducing labor costs, and minimizing fruit damage, all while maintaining high sorting accuracy.

From a perception standpoint, our use of augmented datasets and an RCNN-based object detection pipeline provided promising results despite limitations in the training environment. With few training samples, we achieved mask segmentation and object detection accuracies surpassing 90 percent, demonstrating the strength and adaptability of our data generation and machine learning approach. Though further work is required to scale up the training data and improve the model's performance in more varied environments, these initial results underline the potential for near-perfect accuracy with sufficient computational resources and larger datasets.

On the motion planning front, we employed trajectory optimization techniques within Drake and RRT-based methods to enable safe and collision-free movements. While the results did not meet our initial speed targets, the system successfully guided the arm to designated fruit, validating the desired approach. Future refinements may involve more efficient path planning algorithms, hardware acceleration, and improved optimization strategies, all of which would help meet strict industrial time requirements.

Grasping remains an area for further development. Our preliminary demonstrations using antipodal heuristics and simple point cloud processing show that stable grasps can be achieved on well-defined objects, but integrating deformability considerations and force control into a complete pipeline proved challenging. The inability to process deformable fruit models due to technical difficulties with vtk file formats prevented full validation. Nevertheless, the proof-of-concept illustrates the potential to adapt grasp strategies to different fruit types, shapes, and ripeness levels, reducing damage and preserving quality.

Moving forward, the project could benefit from more extensive training datasets, automated deformability calibration, and improved perception and planning modules. Expanding the fruit library to include visually similar produce and generalizing to new environments would enhance system robustness. The long-term vision is an integrated system capable of running seamlessly on a moving conveyor, autonomously identifying, sorting, and delicately handling an assortment of fruits at production scale. Ultimately, this research marks a significant step toward automation solutions that can revolutionize the agricultural industry, providing greater speed, accuracy, and consistency while protecting produce quality.

Contribution Statement

Kristian did work on the project that pertained to the scene design and simulation creation. It also included motion planning and grasp control. This includes subjects like collision avoidance, grasp control, point cloud processing, and the general diagram/system design.

John did work related to perception and object initialization. This included generating datasets and finding ways to train a perception module to isolate and identify fruit from RBG image feeds. It also included finding models, importing them, and integrating them with the simulation environment, along with configuring deformable entities to work in the simulation.

References

[1] "Manual Harvesting - an overview — ScienceDirect Topics," www.sciencedirect.com. https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/manual-harvesting

[2] S.Abdalla, S.S.Apramian, L. F. Cantley, and M. R. Cullen, "Occupation and Risk for Injuries," PubMed, 2019. https://www.ncbi.nlm.nih.gov/books/NBK525209/

[3] Campos, C.B., et al. (2011). "Technologies for robot grippers in pick and place operations for fresh fruits and vegetables.", INIA, 2011, https://dialnet.unirioja.es/servlet/articulo?codigo=8830140

[4] C. Mitash, A. Boularias, and K. Bekris, "Physics-based scene-level reasoning for object pose estimation in clutter," The International Journal of Robotics Research, p. 027836491984655, May 2019, doi: https://doi.org/10.1177/0278364919846551.