Robust Pick and Pour for Autonomous Plant Watering

Shivangi Gupta[1], Shara Lawande[1], Paramjit Baweja[1], and Qilin Wu[1]

***Abstract*— This report presents an autonomous picking and pouring system using a 6-degree of freedom Franka Emika robotic arm equipped with a 2-jaw gripper and RealSense RGBD camera mounted on the end effector. Our system precisely identifies the circular rims in plastic glasses, effectively grips the cup and pours the contents into an empty glass. This procedure thus acts as a reset for itself. MoveIt's planning capabilities are used to plan collision-free trajectories at each step. Our system design was evaluated on 2 cups placed arbitrarily in the Franka workspace. Our results show a robust system capable of idenitifying filled- and empty- cups, irrespective of colour and contents, with a 80% accuracy. This system is suitable for applications involving pouring, with the primary use case being for autonomous plant watering system.**

I. INTRODUCTION

In the realm of autonomous systems, the ability to perform intricate manipulation tasks such as picking and pouring is essential for advancing robotic applications in everyday scenarios. This report presents the development of an autonomous pick and pour system using a Franka Emika robotic arm, augmented with visual feedback and a gripper. The motivation behind this project comes from a possibility to autonomously water plants, a task which may seem simple yet is technically demanding task and involves accurate object handling and fluid manipulation.

Our approach utilizes OpenCV's HoughCircles for accurate cup detection and a custom algorithm for filled detection based on the variance of the Laplacian, which helps in identifying the filled state of cups. These methods ensure that the robotic arm can recognize and differentiate between filled and empty cups across various environmental conditions. Moreover, the system uses the RealSense RGBD camera for real-time spatial analysis, enhancing the robot's interaction with its environment.

The manipulation is supported by an improved gripping mechanism which has undergone iterations to optimize the pick and pour process. This includes adjustments in the gripping technique and modifications to the pouring strategy to handle different fluid contents effectively. The system uses the MoveIt framework for planning and executing collision-free trajectories, thus ensuring safe and efficient operations within our workspace.

Our testing in diverse settings demonstrates the system's effectiveness in autonomously transferring content from one cup to another, mimicking the process of watering plants. The following sections will elaborate on the relevant past work in this area, the methodology we have followed for each subsystem, evaluation criteria, challenges faced during this system design, and future scope for improvements in our work.

*This work was supported by the Robot Autonomy course project

[1]All the authors are graduate students of the Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA. shivangi,slawande,paramjib,qilinw@cs.cmu.edu

II. RELEVANT WORK

The field of autonomous robotic pouring has quite a lot of incredible prior work, with developments in both the robotic manipulation and visual detection systems. A key aspect of this progress involves the integration of robust vision modules that allow robots to perform precise and adaptive pouring tasks. For instance, using pre-trained segmentation models [1], robots can effectively detect and plan pouring into transparent containers by adapting their actions based on visual analysis of the liquid's state and container's capacity.

Another critical area of development is using robot learning techniques, such as enhancing robotic pouring through hierarchical imitation learning [2]. This approach involves collecting demonstration data and employing a model that supports real-time decision-making in drink pouring tasks.

Robotic pouring can also benefit from advanced control strategies, such as employing auditory and haptic feedback to adjust to the environment and task conditions. This multimodal approach aids in overcoming challenges associated with vision-based methods, especially in occluded or noisy environments. [3] Furthermore, leveraging model predictive control (MPC) enabled by recurrent neural networks (RNNs), robots can achieve high accuracy in pouring tasks, showcasing adaptability to different container types and liquid properties. [4]

These advancements collectively show the state-of-the-art in robotic capabilities, demonstrating significant potential for autonomous manipulation in the coming future. The ongoing research continues to focus on improving the accuracy, adaptability, and explainability of these systems, ensuring their effectiveness across multiple environments.

III. METHODOLOGY

In our project, we aimed to automate the task of identifying and distinguishing between filled and empty cups, precisely planning and moving to the filled cup, grasping the cup and pouring it precisely over the empty cup. This task would then be repeated again, with the manipulator pouring from the newly-filled cup back to the newly-empty cup. In this way, our method served as a rest process for itself and could go back and forth an infinite number of times in case no failure case is encountered.

To build our system, we defined two primary subsystems: the perception pipeline, and the manipulator planning (including the gripper mechanism). Once these subsystems achieved their individual goals, they would be integrated to build the final system. Below, we highlight our methodology for each subsystem, as well as the process for integration.

A. *Perception System*

The perception system's primary goal was to accurately detect and classify the cups based on their status. Initially, we considered using the YOLOv9 object detection model for its robustness in detecting various objects with high accuracy across different environments; however, we shifted towards a more specialized circle detection approach. This decision was driven by specific project needs and constraints, which are outlined below.

- Complexity: YOLOv9 being an intensive object detection model requires substantial computational resources, which were not justified given the simplicity of our task.
- Training Requirements: To achieve optimal performance, YOLOv9 would require training on a significant dataset of images depicting filled and empty cups under various conditions. Generating and annotating this dataset would be too time-consuming and redundant.
- Specificity of Task: Our task was limited to identifying circular objects (cups), which made specialized circle detection algorithms a more efficient choice. These algorithms are optimized for circular pattern recognition, offering faster processing times and lower computational overhead.

The result of this approach are shown in Fig 1.

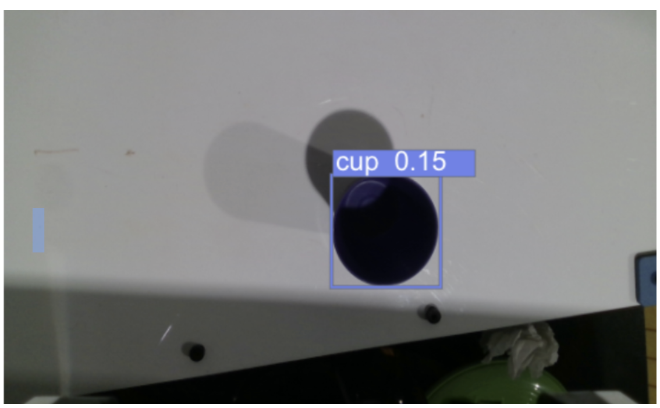

Fig. 1. Filled Cup Detection

Given the project's scope, we adopted a circle detection approach using the Hough Circle Transform, an algorithm suited for detecting circular shapes in images. This method provided a balance of accuracy and efficiency crucial for real-time applications. The methodology involved several key steps:

- Image Preprocessing: Images captured by the onboard camera were first converted to grayscale to reduce computational complexity. This conversion simplifies the data the algorithm needs to process, focusing solely on intensity values which are sufficient for detecting geometric shapes.
- Gaussian Blurring: To reduce noise and improve detection accuracy, a Gaussian blur was applied to the grayscale images. This step helps to smooth out irregularities and enhance the definition of edges.
- Circle Detection with Hough Transform: We utilized the Hough Circle Transform to detect circles. This technique works by transforming points in the image space to a circular parameter space and detecting the presence of circles based on the accumulation of intersecting curves in the parameter space. We adjusted the algorithm's parameters (e.g., resolution, minimum distance between circles, and threshold values) to optimize detection accuracy for our specific application.
- Interactive Verification: To ensure reliability, we implemented an interactive verification step where the detected circles are presented to an operator who confirms or rejects the detection. This step helped to refine the results and ensure the highest level of accuracy for subsequent operations by the robotic arm.
- Variance-Based Classification: Post-verification, the filled and empty cups were classified based on the variance of the Laplacian, calculated within the confirmed circles. This measure helped distinguish between cups with flat surfaces and those containing objects (colored blocks simulating water), as variance is higher where image features such as edges and texture are more pronounced.

The results of this approach are shown in Fig 2.

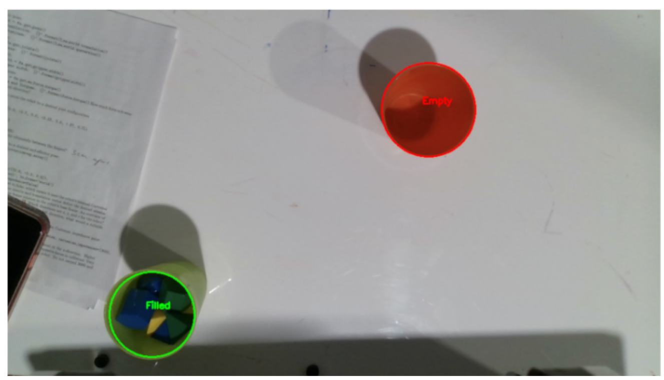

Fig. 2. Filled Cup Detection

B. *Planning and Manipulation*

The planning and manipulation module kicks in once the cups are detected by the perception stack. The system transforms the position to the base frame of the robot, so that the cup can be picked up and poured. The gripping position is on the brim of the cup on the right side. This was determined after a lot of iterations, and is elaborated on later in this section.

These real world coordinates are then fed into the planner, which converts the global pose, from cartesian to joint space, using the Inverse Kinematics solver in the system. Then the planner finds the most optimal set of movements to move from the current joint values to the goal joint values. This is separate from the gripper motion to grab and drop the cup. The planner factors in the joint velocity and extreme limits, to plan accordingly, so that the Franka Arm can track the planned trajectory as is. The plan is stored as list of goal joint angles, and is given in a time-optimised manner (using the joint velocity limits) to the arm to follow.

Coming to the gripper mechanism, there was a lot of iteration required to reach the optimum way to grasp the cup. Initially, the cup was gripped with either jaw across the circumference, as seen in the figure. The gripper was set to a certain width which allowed the cup to be held tightly while the cup was rotated to drop the contents.

Fig. 3. Deprecated gripping mechanism across diameter

This mechanism seemed satisfactory enough, but the pouring pose was not very natural to a typical way a human would pour something out of a cup. Furthermore, there were instances where items would get stuck inside the gripper while pouring, and this method also required the cup estimation to be absolutely perfect for the gripper to hold across the diameter of the cup.

To this end, we pivoted to gripping the cup from just one edge. This method was a lot more robust and would succesfully grip the cup even with some imperfections in the

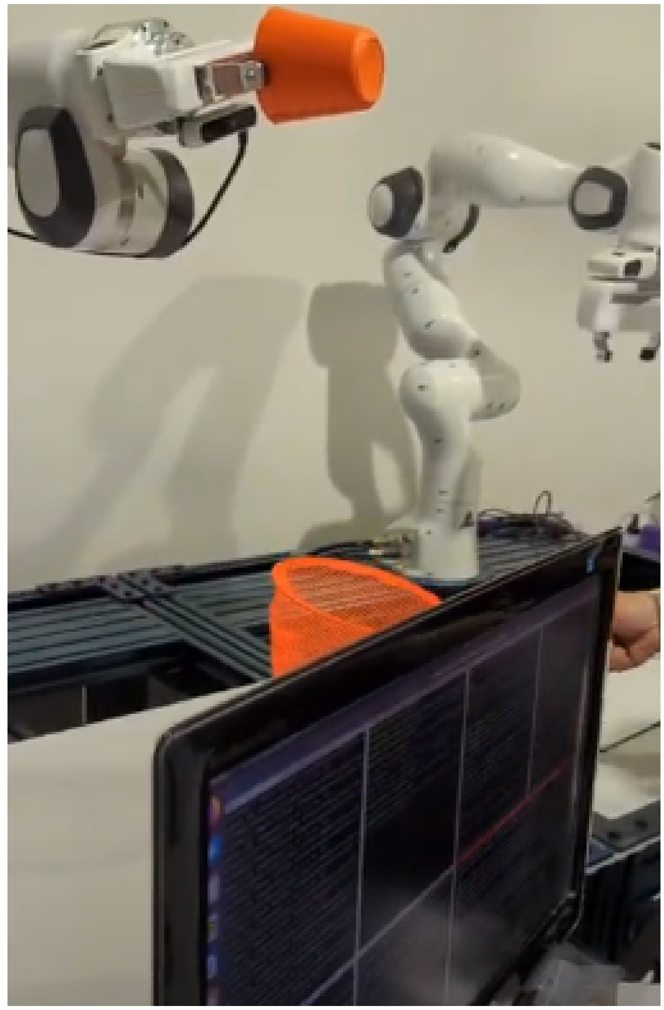

Fig. 4. Deprecated awkward pouring pose

calculated cup coordinates. The cup also remained exposed instead of being covered by the gripper, allowing the contents to be poured out without obstruction. We were also able to set better pouring poses for the manipulator at lower heights, making it a worthwhile decision. This design was finalised for our system.

C. Integration

To begin with, we defined a position such that the entire workspace is in the frame of the camera, as shown in Fig 5. Overall the robot takes a set of steps to complete the task, bringing together the perception and manipulation subsystems. The steps are as follows:

1) **Go to a pre-defined "imaging" pose:** Move the robot arm to a specific position and orientation optimized for capturing images of the cups. This pose should provide a clear view of the cups for subsequent processing.
2) **Find the location of the cups in the image:** Utilize computer vision algorithms to analyze the image captured in the previous step and detect the positions of the cups within the image.
3) **Determine which cup is filled, and which is empty:**

Fig. 5. Stretched position of the arm for the camera.

Analyze the detected cups to determine which one contains liquid (filled cup) and which one is empty. This determination may be based on color, shape, or other visual cues.

4) **Transform the locations from the image frame to the base frame of the robot:** Convert the positions of the detected cups from the camera's reference frame (image frame) to the robot's reference frame (base frame) to enable the robot to interact with the cups accurately.
5) **Calculate the gripper pose to pick the filled cup up:** Determine the optimal position and orientation for the robot's gripper to grasp the filled cup securely without spilling its contents.
6) **Go to a pre-pouring pose over the empty cup:** Move the robot arm to a predefined position and orientation located above the empty cup, ready for pouring liquid into it.
7) **Calculate the pouring pose over the empty cup:** Determine the optimal position and orientation for pouring the liquid from the filled cup into the empty cup while minimizing spillage and ensuring accurate pouring.
8) **Go to the calculated pose:** Move the robot arm to the calculated pouring pose over the empty cup, ready to initiate the pouring process.
9) **Keep the cup back to its initial position:** After pouring the liquid, return the filled cup to its initial position or a designated location to maintain order and prepare for the next cycle of cup filling.
10) **Repeat all these steps:** Once the pouring process is complete, repeat the entire sequence of steps in an infinite loop filling the empty cup.

IV. Evaluation

To comprehensively evaluate the performance of the robotic system, we conducted a series of approximately 15 experimental tests within our designated workspace, which was equipped with both an empty and a filled cup. During each test, the robot was tasked with executing the pick-and-pour operation twice to simulate realistic usage scenarios. Our observations confirmed that the system was capable of consistently identifying the filled cup with an accuracy exceeding 80%. Furthermore, the robot demonstrated flawless precision in localizing and gripping the filled cup, achieving a success rate of 100%. The subsequent pouring of the material into the empty cup was accomplished with a commendable accuracy of 90%. After completing the pouring task, the system consistently returned the cup to its original position with absolute accuracy, recorded at 100%.

In instances where either the picking or the pouring process deviated from the expected accuracy, manual intervention was necessary to rectify the issue and restart the system. This step was essential to ensure the continuity and reliability of the operational testing, providing valuable insights into the system's capabilities and areas requiring further enhancement

Fig. 6. Gripping the Filled Cup

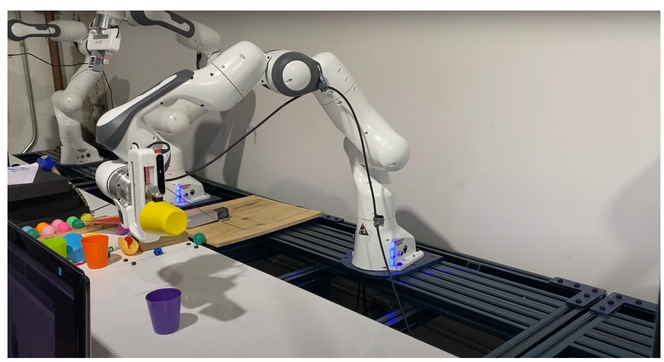

Fig. 7. Pouring Material into Empty Cup

TABLE I
RESULTS

Total Runs	15
Filled Cup Identification	80%
Grip Filled Cup	100%
Pour Material	90%
Place Cup to Original Position	100%

In the analysis presented, it becomes evident through Figure 2 that the system is capable of accurately distinguishing

Fig. 8. Placing Filled Cup Back to it's Original Location

between filled and empty cups. Following this identification phase, the mechanism successfully secures the filled cup, as depicted in Figure 6. Subsequently, the contents of the filled cup are meticulously transferred into an empty cup, a process clearly illustrated in Figure 7. Finally, as shown in Figure 8, the now-empty original cup is precisely repositioned back to its initial location, completing the cycle of operations.

V. Challenges

A. Perception System

During the development of our perception system, we encountered several technical challenges that required innovative solutions and adjustments to our initial approach.

- Lighting Variability: One of the primary challenges in visual perception systems is dealing with variable lighting conditions. Changes in ambient light affected the accuracy of circle detection algorithms, leading to inconsistent results. To address this, we implemented adaptive thresholding within our preprocessing pipeline, allowing the system to maintain performance across different lighting scenarios by dynamically adjusting detection parameters based on the observed lighting conditions.
- Circle Detection Accuracy: Initially, our system struggled with accurately distinguishing between filled and empty cups due to similar shapes and overlapping features. The Hough Circle Transform sometimes misidentified non-circular objects or missed circles altogether. To mitigate these issues, we refined the parameters of the Gaussian blur and the Hough Transform based on extensive testing. Additionally, we introduced an interactive verification step, where a human operator could confirm the accuracy of detected circles, significantly enhancing the reliability of our detections.
- False Positives in Object Recognition: Early tests revealed a high rate of false positives, where objects other than cups were being mistakenly identified as target objects. We tackled this challenge by integrating a variance-based method to analyze the texture within the detected circles, which helped distinguish the flat surface of empty cups from the textured surface of filled cups, marked by colored blocks simulating water.

B. Manipulation System

On the manipulation side, our team faced distinct challenges primarily related to the physical interaction of the robotic arm with the environment:

- Pouring Angle Precision: Setting the correct angle for pouring presented a significant challenge. The robotic arm needed to be precise not only in positioning but also in the orientation when performing the pouring action. Achieving this required numerous trials and adjustments. Through iterative testing, we developed a series of calibrated movements that allowed the robotic arm to align accurately with the target container, ensuring successful pouring without spillage.

VI. Future Work

In envisioning the future enhancements of our robotic system, we aim to integrate more advanced models that allow for end-to-end self-controlled learning. This approach would enable the robot to independently understand and adapt to various tasks through natural language prompts and visual cues, thereby achieving higher autonomy and flexibility in operation. There are some works that we can take inspiration from like.

- Advanced Grasp-Action-Target Learning Recent studies, such as "GATER: Learning Grasp-Action-Target Embeddings and Relations for Task-Specific Grasping" (IEEE RA-L, 2021), highlight significant progress in learning embeddings that specifically cater to task-oriented grasping scenarios. This research presents a method where robots learn to associate specific grasping actions with particular objects and scenarios, improving efficiency and adaptability in real-world tasks.
- Few-Shot Language-Guided Manipulation Another promising area is the development of models capable of few-shot, language-guided manipulation as explored in "Distilled Feature Fields Enable Few-Shot Language-Guided Manipulation" (CoRL 2023). These methods utilize distilled knowledge from large datasets to perform specific tasks with minimal direct instruction, enabling robots to understand and execute tasks from limited examples and simple language descriptions.
- Integration of Deep Learning and Robotics Further into the future, as discussed in "Shaping the future of advanced robotics" by Google DeepMind (2024), there is potential for merging deep learning techniques with robotic control systems to create highly intuitive and adaptive robotic assistants. This approach could

revolutionize how robots interact with complex environments and perform a variety of tasks based on dynamic learning and decision-making processes modeled on human-like cognition.

VII. Conclusion

Our project successfully developed an autonomous robotic system for precision pick-and-pour tasks, a critical advancement for applications such as autonomous plant watering. Utilizing a 6-degree of freedom Franka Emika robotic arm equipped with a RealSense RGBD camera and an optimized 2-jaw gripper, the system demonstrated robust capabilities in identifying and manipulating cups within a static workspace.

To detect the cups accurately we integrated the 'HoughCircles' method provided by OpenCV for initial detections and enhanced object recognition through average variance calculations of the Laplacian.

In terms of manipulation, the Franka Arm was programmed to execute tasks of precise picking and controlled pouring of materials based on the detected filled and empty cup locations. The initial challenges with grip stability and awkward pouring dynamics were addressed by redesigning the gripping mechanism, which significantly improved the system's performance. Throughout our series of evaluations involving 15 experimental tests, the system reliably identified filled and empty cups with high accuracy, executed precise grips, and effectively poured contents between cups. The accuracy rates achieved—80% in cup identification, 100% in gripping and repositioning, and 90% in pouring.

Future work could involve the integration of advanced object detection models, such as the Grounding-DINO, and language-guided manipulation strategies to create more adaptable and intelligent robotic systems. These improvements are expected to not only refine the system's operational efficacy but also broaden its application spectrum to more complex and nuanced tasks.

Acknowledgment

We would like to take this opportunity to thank Dr. Oliver Kroemer for his guidance and suggestions while we were defining our problem statement, and during the mid semester check-in. We would also like to thank our teaching assistants, Abhinav and Vibhakar, for their valuable insights and assistance in debugging errors faced while operating the manipulator. A huge thank you also to our wonderful MRSD cohort for their insights and companionship throughout this project.

References

[1] F. Zhu, S. Hu, L. Leng, A. Bartsch, A. George, and A. B. Farimani, "Pour me a drink: Robotic precision pouring carbonated beverages into transparent containers," 2023.

[2] D. Zhang, Y. Zheng, Q. Li, L. Wei, D. Zhang, and Z. Zhang, "Explainable hierarchical imitation learning for robotic drink pouring," 2021.

[3] H. Liang, C. Zhou, S. Li, X. Ma, N. Hendrich, T. Gerkmann, F. Sun, M. Stoffel, and J. Zhang, "Robust robotic pouring using audition and haptics," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Oct. 2020.

[4] Y. Huang and Y. Sun, "Accurate robotic pouring for serving drinks," 2019.