Winning the Amazon Picking Challenge 2016: Team Delft System Overview and Lessons Learned

Mukunda Bharatheesha\textsuperscript{1}, Carlos Hernández Corbato\textsuperscript{1} and Martijn Wisse\textsuperscript{1}

\textbf{Abstract}—Team Delft’s robotic system won both the Picking and Stowing Competitions at the Amazon Picking Challenge 2016. The goal of the challenge is to automate pick and place operations in unstructured environments, specifically the (representative) shelves of an Amazon warehouse. Team Delft’s robot is based on an industrial robot arm, 3D cameras and a customized gripper. The robot’s software uses ROS to integrate off-the-shelf components and modules developed specifically for the competition such as implementing Deep Learning and other AI techniques for object recognition and pose estimation, grasp planning and motion planning. This paper provides an overview of the main functional components of the system, and discusses their performance and results at the Amazon Picking Challenge 2016 finals.

I. INTRODUCTION

The Amazon Picking Challenge 2016 included two competitions: the Picking and Stowing competitions. In the Picking competition 12 items from the competition product set had to be picked from an Amazon shelving unit and placed in a tote; in the Stowing competition it was the other way around: 12 items were to be picked from the tote and stowed into the shelf. The maximum allotted time to fulfil each task was 15 minutes and the system had to operate autonomously. A file containing the task order was given to the system, which included the initial contents of the bins and the tote, and it had to produce a resulting file indicating the location of all the products.

The set of 39 items used in the challenge were representative of those in an Amazon warehouse. Books, cubic boxes, clothing, soft objects, and irregularly shaped objects represented realistic challenges such as reflective packaging, different sizes or deformable shapes. The items could be placed in any orientation inside the bins, sometimes cluttering them, and the target product could be partially occluded by others.

Teams had to place their robots in a $2 \times 2$ m workcell, no closer than 10 cm from the shelf. The workspace also posed important challenges to perception and manipulation. The shelf was a metal and cardboard structure divided into a matrix of $4 \times 3$ bins. The bins were narrow but deep, which limited the manoeuvrability inside and required a long reach. Additionally, the shelf construction resulted in significant deviations in reality from its ideal geometric model.

The performance of the robots during the picking and the stowing tasks was evaluated by giving points for correctly placed items and subtracting penalty points for dropping, damaging or misplacing items. A correct operation could receive 10, 15 or 20 points depending on the cluttering of the bin. Additional bonus points were given for specially difficult objects, for maximum scoring of 185 points in the Picking Task and 246 points in the Stowing Task.

II. ROBOT SYSTEM DESCRIPTION

Team Delft’s robotic system is based on an industrial robot arm, a custom made gripper and 3D cameras, as shown in figure \[3\]. For the robot arm we chose a 7 DOF SIA20F Motoman mounted on a horizontal rail perpendicular to the shelf. The resulting 8 degrees of freedom allowed the system to reach all the bins with enough manoeuvrability to pick the target objects as indicated in figure \[2\].

We customized our own gripper to handle all the products in the competition (see figure \[3\]). It has a lean footprint...
to manoeuvre inside the bins, and a 40cm length to reach objects at the back. It includes a high flow suction cup at the end, with a 90 degrees rotation allowing two orientations, and a pinch mechanism for the products difficult to suck. Both the suction cup rotation and the pinch mechanism are pneumatically actuated. A vacuum sensor provides boolean feedback whether the suction cup holds anything. For object detection a 3D camera is mounted in the gripper to scan the bins, while another one is fixed on a pole above the tote. The tote is placed on a frame attached to the robot rail. The compressor and the vacuum pump required to actuate the gripper are mounted on another frame that attached to the rail base, so the whole set up could be easily moved in three big blocks. Robust and easy transportation and installation were important requirements.

The system control is based on the sense-plan-act paradigm and path planning for robot motion. This allows for potentially optimal motions, at the cost of more precise sensing information. First, the task is decomposed into a set of pick and place operations on the target items. Then, for each operation in the Picking task the sense-plan-act cycle proceeds as follows:\footnote{A video demonstrating the pipeline can be found here: \url{https://www.youtube.com/watch?v=PKgFy6VUC-k}}

In the sense step the robot moves to take an image of the bin containing the first target item to locate it and get the obstacles information. Then, during the plan step a grasping strategy and candidate pose for the gripper to grab it are computed, and a motion plan is generated to approach, grasp and retreat from the bin with the item. Finally, in the act step the gripper is configured for the selected strategy and the complete motion is executed, including gripper activation to suck or pinch-grasp the item.

The vacuum seal in the suction cup is checked to confirm a successful pick. If so, the robot moves to deposit the item in the tote, using simple drop-off motions. This cycle is repeated till all target items are picked. For the Stowing task the loop operates similarly until all items in the tote are stowed in the shelf.

III. ROBOT SOFTWARE

Team Delft was fully committed to the ROS-Industrial initiative [1] that aims to create industry-ready, advanced software components to extend the capabilities of factory robots. The robot software is thus based on the ROS framework [2]. We found that the flexibility, modularity and tools provided by ROS allowed us to address the requirements for autonomy and high and reliable performance in the competition, and facilitated development.

The ROS component-based approach allowed for the integration of the different components for task management, object detection, pose estimation, grasping and motion planning into a robust architecture. Following we describe them.

On top of the architecture sits the task manager, responsible for decomposing the Pick and the Stow tasks into a plan of pick and place operations, and manages the state of fulfilment of the whole task. It encodes the competition rules to maximize the scoring, by planning first those operations that scored more points, and keeps track of the location of all the items.

A central coordinator module coordinates the execution of each pick and place operation following a sequential flow. It was implemented as a ROS SMACH [3] state machine.

The system can handle some failures applying fallback mechanisms to continue operation. For example, if the robot cannot find the target item, or estimate its pose, it tries different camera viewpoints, then if the problem persists it postpones that target and moves to the next operation. The system can detect if a suction grasp failed by checking the vacuum sealing after execution of the complete grasp and retreat action. If there is no seal the robot assumes the item dropped inside the bin and retries the pick later. If the seal is broken during the placement in the tote, the item is assumed to have dropped in the tote.

The key software modules that are central to the Team Delft robotic system are the vision, manipulation (grasp synthesis) and motion (planning and execution).
each of these modules in detail would be out of the scope of this paper. Hence, we limit the information to a few key aspects and proceed directly to the results achieved at the competition.

The vision module performed the tasks of object detection (including recognition) and estimating the pose of the detected object so that a grasp can be synthesized subsequently. For object detection, a deep neural network using Faster R-CNN [4] is used that provides a bounding box around the object of interest. The corresponding point cloud from the bounding box is used to estimate the pose of the object in the bin using Super 4PCS [5]. Custom 3D-Models of the objects are used as reference inputs as desired by the Super 4PCS algorithm.

The estimated pose of the object is then used by the grasp synthesizer to translate the object pose to a grasp pose for the robot. The grasp synthesizer uses offline geometric constraints and user-defined heuristics are used to synthesize grasp candidates. The fundamental idea is to generate a set of grasp candidates over the surface of the 3D model of the item based on primitive shapes (cylinders, spheres, cones, planes, etc.) and prune them online using geometry constraints due to the actual item’s estimated pose while accounting for possible occlusions.

Finally, the robot motion module generated the required motions to the selected grasp pose for the object of interest while avoiding collisions with the environment. The main strategy of Team Delft’s motion module was to limit the requirement of (the computationally intensive) online planning to only where required and exploit the benefits of working in a static environment. This was achieved by dividing the motions into coarse motions (offline) and fine motions (online). The coarse motions were offline computed joint space trajectories between a defined pair of robot configurations and the fine motions were cartesian space motions to manipulate the object of interest. The entire motion module was developed using multiple APIs from the MoveIt! motion planning library [6]. For further details, we refer to [7].

IV. RESULTS

Team Delft’s robot was the champion of the challenge winning both competitions, with an outstanding performance in the Stowing Task.²

A key focus during testing of our system was the minimization of cycle time. This eventually proved to be Team Delft’s winning formula, particularly in the picking competition where we scored the same points as the runner-up Team PFN. We were declared winners as we achieved our first successful pick in 30 s. In the stowing competition too, the lower cycle times enabled us to finish the competition within half the allotted time. It is pertinent to highlight that the deep learning based object detection framework implemented on state of the art NVIDIA GPUs[8] not only performed robustly, but also rapidly sped up the cycle time. Thus, we could ensure the robot would remain stationary for as less time as possible. However, we do realize that the cycle time could be further reduced by parallelization of the sense and plan stages while act is being executed. With the robotic setup that was presented in Section II, this parallelization could have perhaps been possible for the stowing competition, but we did not have enough development time for the same.

The final scores of the top four teams for the Amazon Picking Challenge 2016 Pick and Stow competitions are shown in Table 1. Apart from faster cycle times, the overall results of the teams improved considerably over the previous APC edition (APC 2015), albeit with an introduction of a new competition (stowing) and with more cluttered bins for the picking competition.

V. LESSONS LEARNED

Considering the results described in the previous section and the complete experience developing the robot for the Amazon Picking Challenge, we reached several conclusions about our concept design premises and how to improve it.

The most important idea is that manipulation requires contact with the environment. Team Delft’s pure planning approach to grasping and manipulation treated contact as collisions to avoid, and simply by-passed this constraint for the target object. This caused a lot of rejected plans to grasp items from cluttered bins, some of them becoming actually unrealisable. Force-feedback and compliance in the gripper seem unavoidable to achieve a reliable solution. Also, creating a single gripper capable of handling such a variety of products proved difficult. None of the teams managed to pick the dumbbell, for example. Having different grippers and switching between them on the fly seems a more efficient and robust solution.

On the perception side, Deep Learning neural networks proved an excellent solution for object recognition, but they also are a really promising solution to pose estimation and even grasp planning, as the results of other teams suggest.

Notwithstanding the discussed improvements, Team Delft’s concept based on speed and reliability proved successful. The ready-for-industry approach we took, with installation and setup procedures, and professional team coordination during the competition, allowed to keep robustly improving the robot’s performance till reaching close to its top limit right at the competition.

²Video recordings of Team Delft’s competition runs can be found here [https://youtu.be/3KlzVxomOq](picking) and here [https://youtu.be/AHuUdVdMfg](stowing)
VI. Conclusions

The overall high scores by many teams, and the excellent performance of Team Delft’s robot, suggest that the bin picking problem for diverse, medium-size products can be addressed by current robotic technology. Speed is still far from human performance (~100 items an hour, compared to 400 items an hour in the case of a human), but considering that Team Delft’s robot could have been speed-up probably 50% with faster motions and faster processing, we are confident to predict that robot technology is getting there. However, the Picking task results, proved that general manipulation, including diverse objects and cluttered spaces, still remains an open problem for robotics.

ACKNOWLEDGMENT

All authors gratefully acknowledge the financial support by the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 609206. We would like to thank RoboValley[3] the ROS-Industrial consortium, our sponsors Yaskawa, IDS, Phaer, Ikbenstil and Induvac, the people at the Delft Center for Systems and Control and TU Delft Logistics for their support, also Lacquey B.V. for helping us handle our heavy rail, and a special thanks to Gijs van der Hoorn of the TU Delft Robotics Institute for his help during the development of the robotic system. Finally, the wonderful team who, along with making us the double champions, also made the whole experience a lot of fun! Wilson Ko, Hans Gaiser, Jethro Tan, Kanter van Deurzen, Maarten de Vries, Bas Van Mil, Jeff van Egmond, Ruben Burger, Mihai Morariu, Jihong Ju, Xander Gerrmann, Ronald Ensing and Jan Van Frankenhuyzen, thank you once again.

REFERENCES