Doctoral Thesis

Novel Performance Metrics to Evaluate Bin-picking Robots for Various Mixed Items

FUJITA Masahiro
Program of Information Science and Engineering
Graduate School of Science and Technology
Nara Institute of Science and Technology

Supervisor: Professor Tsukasa Ogasawara
Robotics Lab (Division of Information Science)

Submitted on October 29, 2020

A Doctoral Thesis
submitted to the Graduate School of Science and Technology,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

FUJITA Masahiro

Thesis Committee:

Professor Tsukasa Ogasawara	(Supervisor)
Professor Kenji Sugimoto	(Co-supervisor)
Associate Professor Jun Takamatsu	(Co-supervisor)
Assistant Professor Gustavo Alfonso Garcia Ricardez	(Co-supervisor)

Novel Performance Metrics to Evaluate Bin-picking Robots for Various Mixed Items*

FUJITA Masahiro

Abstract

Bin picking of various mixed items is an important research problem in robotics. Recent robotics competitions are an excellent platform for technology comparisons since some participants may use state-of-the-art technologies, while others may use conventional ones. Nevertheless, even though points are awarded or subtracted based on the performance in the frame of the competition rules, the final score does not directly reflect the suitability of the technology used. Therefore, it is difficult to understand from the scores which technologies and their combination are optimal for various real-world problems. One key element for a bin-picking robot is the gripper, as it is the main tool to manipulate objects. If the target items are diverse, multiple grippers are normally used. A design of gripper combinations depends not only on the item variations but also on the state of the bins, which changes with the progress of the task.

This dissertation first proposes a strategy to change the gripper combination during a bin-picking task based on the sparseness of objects inside bins, and a bin-picking robot system using it. The evaluation results using successful picking rate as a metric are shown, and the effectiveness of this strategy is verified. However, this metric represents only one aspect of the robot system as well as scores in the competitions.

Furthermore, this dissertation proposes a set of performance metrics selected in terms of actual field use as a solution to clarify the important technologies in bin picking. Moreover, we use the selected metrics to compare four original robot systems, which achieved the best performance in the Stow task of the Amazon Robotics Challenge 2017. Based on this comparison, we discuss which technologies are important for practical use in bin-picking robots in the fields of warehouse automation.

*Doctoral Thesis, Graduate School of Science and Technology, Nara Institute of Science and Technology, October 29, 2020.

Keywords:

bin picking; various items; gripper combination; performance metrics; warehouse automation

多品種混載ビンピッキングロボットのための性能評価指標群*

藤田 正弘

内容梗概

多品種混載物のビンピッキングはロボット工学における重要な研究課題である。近年のロボット競技会は、最新技術を用いる参加者と従来技術を用いる参加者がいるため、技術比較の優れたプラットフォームであると言える。競技ルールに基づいて加点、減点が行われるが、最終的なスコアが用いられた技術の適合性を直接的には反映しないため、スコアからどの技術とその組み合わせが実世界の様々な課題に対して最適であるかを理解することは困難である。また、ビンピッキングロボットの重要な要素の一つがグリッパであり、対象物が多様な場合には、通常、複数のグリッパが用いられる。複数のグリッパの組み合せ設計は、対象物のバリエーションと共に、作業の進行に応じて変化する容器内の状態にも依存する。

本論文では、まず、容器内における「物体のまばらさ」に基づいたグリッパの組み合せ戦略と、それを用いたビンピッキングロボットシステムを提案する。そして、把持成功率を指標とした評価結果を示し、本戦略の有効性を検証する。ただし、この指標は競技会のスコアと同様にロボットシステムのある一つの側面しか表していない。

そこで、次に、ビンピッキングにおける重要な技術を明らかにするために、実際の現場使用の観点から選択した新たな性能評価指標群を提案する。そしてAmazon Robotics Challenge 2017 の Stow タスクにおける上位 4 チームのロボットシステムをこの性能指標群を用いて比較し、物流倉庫自動化の実現に向け、ビンピッキングロボットに重要な技術を論じる。

キーワード

ビンピッキング, 多品種混載, グリッパ組み合わせ, 性能指標、物流倉庫自動化

*奈良先端科学技術大学院大学 先端科学技術研究科 博士論文, 令和 2 年 10 月 29 日.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Bin picking is still an important problem in robotics. Its difficulty is described in [1], for example. Picking an item from a cluttered scene is applied to various fields: parts supply in Factory Automation (FA), pick-and-place in Warehouse Automation (WA), cleaning up using household robots and so on. But it is difficult to apply existing methods, if the target items have various shapes, and materials.

These last few years, robotic competitions which aim to solve a domain challenges are often held as a platform to accelerate technology development. In the FA field, for example, the National Institute of Standards and Technology of USA is organizing the Agile Robotics for Industrial Automation Competition [2] since 2017, which is a simulation-based competition focused on agility. The Ministry of Economy, Trade and Industry of Japan is carrying out an Assembly Challenge [3, 4] in their World Robot Summit[1] since 2018. This is a real robot's competition focused on factors during setup changes and those of during operation. The former factors are agility and leanness. The latter ones are operation rate improvements.

In the WA field, Amazon Robotics, Inc. held a competition in 2017 regarding a warehouse task automation applying robotics. In large warehouses of e-commerce corporations, a mixture of daily items are manually picked and placed. The automation of the manual work is an important problem in robotic bin-picking. In particular, technical problems lie in picking items with various shapes and in identifying their texture, shape, and material. Various methods have been proposed to solve the problems in the competition in which the ability of the robot system was tested in a competition setting.

In this dissertation, we show a system comparison of the four unique teams which ranked first to fourth places in the Stow task of the Amazon Robotics Challenge 2017. In the competition, the systems were ranked according to the organizer's rules. However, it is difficult to analyze the system performance in terms of a more practical use. The reason is that a single metric like the competition's score represents only one as-

[1] https://worldrobotsummit.org/en/

pect of the robot system and loses some information about the original data that reflects its performance. While the needs of each industry are different, competitions reflecting the needs of each industry are taking place. In this dissertation, we first propose a strategy to change the gripper combination during a bin-picking task and a robot system using it. Then, we propose a set of metrics in order to reflect the detailed behavior of a system to our system performance analysis as a solution to clarify the important technologies in bin-picking. Finally, we describe an example analysis to show the issues that arise when robots are applied to actual warehouses by using the comparison results. Also, the elements in each system to be improved to get better performance are clarified.

The main contributions of this dissertation are the following:

- A proposed strategy to change the gripper combination during a bin-picking task and its application to a real robot system.
- Details of four unique robot systems and a comparison of the system configurations of each team.
- A proposed system performance evaluation based on plural metrics introduced from reliability engineering [5, 6]. By this analysis, each team strategy was revealed.
- A discussion on pros and cons of the systems and technologies including the details of the systems, and also a discussion on lessons learned and on future system design.

This dissertation is organized as follows. Chapter 2 introduces the problem setting of the Stow task. Chapter 3 presents the related works. Chapter 4 presents our proposed multi-gripper switching strategy and our proposed robot system using it. Chapter 5 presents our proposed system performance evaluation metrics, describes four systems developed for this task, and presents a performance evaluation using the proposed metrics. Chapter 6 analyzes our findings and presents lessons learned. Finally, Chapter 7 concludes this dissertation.

Chapter 2

Problem setting

Mitsubishi Electric Corporation, Chubu University and Chukyo University participated in the Amazon Robotics Challenge 2017 as team MC^2 with a robot system which implements the proposed gripper designs, gripper combination strategy, and recognition methods. Our robot system picked 18 out of 20 items in the aforementioned Stow task and came out third.

We show an outline of the Stow task rules in the Amazon Robotics Challenge 2017.

- Robots must pick various daily items automatically from a tote and place them into a storage.
- Items are mixed in a tote as shown in Fig. 1. Robots must identify the items and record which items are placed into a bin in the storage.
- If a robot can pick and classify an item, then place the item into a bin in the storage and record its location successfully, points are awarded.
- If items are dropped or damaged while picking and placing, points are subtracted.

Figure 1: Example of items in a tote for the Stow task.

- 20 items are in a tote. Half of the items are distributed to teams in advance. The other half of the items are distributed just before the starting the competition. The recognition dataset must be updated in a few minutes.
- Robots must finish the task in 15 minutes. If robots finish the task before the 15-minute period, teams get additional points.
- Storage can be designed by each team with some limitations: size, number of bins and so on.

We also describe the details of the scoring system in the rules. Points are awarded as follows:

- 5 points for each item stowed into the storage system, plus 5 additional points if the item is a new (unknown) item.
- 1 point for every 5 seconds or fraction thereof that remain on the clock when the task is complete and all items have been successfully stowed in the storage system, so long as at least 15 of the locations of the items are correctly registered in the item location file.

Points are subtracted as follows:

- -15 points for each item that is not in the storage system, stow tote, or amnesty tote at the end of the task, except for items grasped by the robot under normal operation when time runs out.
- -5 points for each item in the storage system or stow tote with an incorrect final location in the item location file.
- -5 points for any item that is dropped into the storage system from a height of more than 15 cm.
- -5 points for each item that is protruding more than 2 cm out of the storage system.
- -5 points for minor damage to an item, such as bends and dents.

- -20 points for major damage to an item, such as large rips, holes, or crushing.

More details, please check the official site[2].

[2]https://www.amazonrobotics.com/site/binaries/content/assets/amazonrobotics/arc/2017-amazon-robotics-challenge-rules-v3.pdf

Chapter 3

Related works

The related works are presented in five groups. The first group includes systems developed for the bin-picking tasks. The second group summarizes different design approaches for grippers. The third group is on grasp planning. The fourth group is on item recognition. Finally, the fifth group is about comparison analysis of competition systems.

3.1 Bin-picking systems related to competitions

Bin-picking is a classical but still state-of-the-art challenge in robotics. Many proposals were made on the automation of pick-and-place tasks in warehouses in the Amazon Picking Challenge (later known as Amazon Robotics Challenge), held from 2015 to 2017. In particular, there were many proposals and findings on gripper design to solve problems when picking various items as described in later sections. In the first competition held in 2015, actual shelves used in Amazon warehouses were also used in the pick-and-place of daily items from a bin.

As mentioned in the summary article of the 2015 competition [7], it was proven that a suction gripper is able to pick many kinds of items. Further in the 2016 competition, the winner Delft [8] and many other teams succeeded in picking hard-to-pick items such as a mesh cup, which a suction gripper was unable to handle, by combining suction and two-finger grasping or similar pinching mechanisms.

In the 2015 competition, we proposed a robot system which can switch between 3 types of two-finger grippers with different widths [9]. In the 2016 competition, we proposed a robot system with a suction gripper and a two-finger gripper [10].

3.2 Gripper design

For grasping, the combination of suction and two-finger or suction only became the common configuration. In the 2017 competition, almost all the teams used either of the two-gripper configurations mentioned above.

The overall winner of 2017 [11] as well as team NAIST-Panasonic (Garcia *et al.* [12]), who configured the system by analyzing the past competitions and came fourth at their first attempt, adopted the gripper configuration of suction and two-finger combination. Team MIT-Princeton [13], the winner of Stow task in 2017, where only bin-picking ability was tested, made a system which enabled several motions such as suction down, suction side, grasp down, flush grasp in one gripper system that combined suction and two-finger grippers. Only the runner-up in Stow task, Team Nanyang [14], used a configuration with two suction grippers and without a two-finger gripper. They achieved a high score by focusing on bin rather than gripper design. To explain in detail, they added a mechanism which expanded the bin space thereby modifying the problem from a hard bin-picking to a simpler picking like a pick-and-place problem from a wide open flat space. The team was successful in picking various items, and their strategy was necessary for items in a hard-to-pick pose or occluded. Team MC^2 proposed a two-stage strategy to use three types of gripper properly [15], which is described in Chapter 4. As aforementioned, many types of gripper designs were proposed in the competitions.

In comparison, jamming gripper [16] is highly versatile in picking various kinds of items. Nevertheless, there are some difficulties in applying jamming grippers to bin-picking because when a jamming gripper tends to pick several small items in a tightly packed bin simultaneously. Also, in principle, as it needs to come into contact with the item before it starts grasping, it tends to fail in picking soft items which may change shape easily. Thus, nobody used the jamming gripper in the competitions.

3.3 Grasp planning and grasp point detection

Grasp point detection takes place to determine the gripper pose to pick detected items. A method [17] which convolutes a binary image model of the gripper to the depth image and does not require pre-information of the object is already used in factory automation. Many methods for grasp point detection using machine learning from RGB images and depth images have been proposed. Jiang *et al.* proposed a method [18] which searches in an RGB image for a pose that is easy for a two-finger gripper to grasp and they were the first to make it practical [19] with deep learning. Pinto *et al.* proposed a grasp point detection method [20] from an RGB image based on 50,000 trials on actual robots. Moreover, Levine and others achieved a method [21] where

hand-eye coordination detects a grasp point from RGB data. The common among all these methods is that they are able to determine the grasp point using only images. Unfortunately, the physical correspondence between gripper and item in grasping [22] cannot be understood just from appearances in an image.

Mahler and others have defined a matrix which determines grasp points for a number of objects in advance from the relationship between the 3D object model and the 3D gripper model and propose a method [23] which assumes a physical grasp point of unknown objects by learning from a vast amount of data. They achieved this with deep learning [24] and used it with vacuum and suction[3] type of grippers [25]. In this method, bin picking is available based on the learned results when the learning becomes precise enough.

Furthermore, Matsumura and others succeeded in bin picking with real robots exclusively by learning from simulation data [26]. Team MIT-Princeton [13] and others fitted for both suction and two-finger grippers by detecting grasp points with Fully Convolutional Network (FCN) on the base [13]. Whether to use a learning method by providing data beforehand or to use a non-learning method which is more adaptive to unknown objects and environmental changes depends on the preconditions of the problem.

3.4 Sensors and algorithms for item recognition

In the competition, item detection based on images obtained from RGB-D sensors is often used. The winning team in 2015 probabilistically classified multi-class items with a method which describes the type of item in each pixel using image features obtained from RGB and depth images [27]. Then, items are segmented by integrating the result. In the 2015 competition, many teams used algorithms based on image features.

From 2016 onward, many adopted Convolutional Neural Networks (CNN) and showed good performance. Faster network variants such as Faster R-CNN [28] which performs bounding box detection and multi-class classification in order, and high speed YOLO [29] and SSD [30], which perform the detection and classification in parallel, were used for the recognition. There were teams [11] who used semantic segmentation methods as a base and all performed well in detecting (classifying) items in the bins.

[3] In this dissertation, we refer to the blower-based suction as *suction* and vacuum-pump-based suction as *vacuum*.

Object pose detection is used to determine grasp point on items after they are detected. In general, data obtained from an RGB-D sensor and object model are matched together using methods such as Iterative Closest Point [31], which minimizes the point cloud position errors between data and model, and Directional Chamfer Matching [32], which presumes object pose by featuring image edges and matching them for every view direction. Such methods are used for bin-picking in factory automation as they are robust against illumination changes, among other things.

The methods [33, 34], which estimate object pose by voting after extracting features and matching pairs of vectors obtained from edges and planes of object, are good in speed and accuracy balance. A method [35], which assumes the position and approximate pose of an object by adding multi-view data to a CNN, is also proposed. Nevertheless, to use these methods, a 3D model of the object is required. If a 3D model is not available, a method by the Team MIT-Princeton [36], which assumes the object pose by fitting primitive shapes like spheres and cubes directly to the data, is also proposed. The method to use depends on the precondition of the problems.

3.5 Comparative analysis of competition systems

In the Amazon Robotics Challenge, points are awarded or subtracted based on the performance in the frame of the rules. But the final score does not seem to directly reflect the technology suitability. Therefore, organizers and teams published papers about system analysis [7]. Results of the analysis are not based on the scores in the competition, but statistical data by a questionnaire survey about used technologies, team configurations, and so on. From the results, we can understand which technologies were well used in the competition. But we have difficulty in analysis of a system performance in terms of more practical use.

Successful picking rate as shown in [17, 15] is an important metric to evaluate the system performance in practical use. Mean Picks Per Hour (MPPH) [23] is also a well-used metric [25, 13] which is related to both picking rate and picking time. But such a single metric represents only one aspect of the robot system and loses some information of the original performance data. Therefore, in this dissertation, we propose a set of metrics in order to analyze a system performance that reflects different aspects of a system behavior.

Chapter 4

Multi-gripper switching strategy based on object sparseness

In large warehouses of e-commerce corporations, a mixture of daily items are manually picked and placed. The automation of the manual work is the most significant problem in robotic bin picking. In particular, technical problems lie in picking items with various shapes and in identifying their texture, shape, and material. Various methods have been proposed to solve the problems proposed in competitions held by Amazon Robotics, Inc. from 2015 to 2017, in which the ability of the robot system was tested in a competition setting. Many teams use a gripper combination of two grippers, such as two-finger and suction, but the combination of grippers may be changed depending on the state of the bins. Therefore, the gripper combination should be adjusted according to such state.

In this chapter, we propose a gripper combination strategy based on sparseness of objects inside bins to change the gripper combination dynamically. In addition, we also propose a robot system which has three different types of grippers and the strategy with which specific items are efficiently picked from a bin containing a pile of various daily items. Furthermore, we describe an item recognition method based on computer vision and its capabilities.

The main contributions of this chapter are the following:

- A proposed metric to switch a gripper combination, determine the object sparseness, and its application to a real robot system
- A robot system for bin-picking based on a combination strategy of three types of grippers.
- A comparison of the successful picking rate between the proposed system and the winner of the Stow task in the Amazon Robotics Challenge 2017.

4.1 Object sparseness

If robots have multiple grippers, the gripper combination should be changed according to the state of the bin, as this state changes when the robot picks items from the bin. At an early stage, many types of items appear on the surface. Therefore, any type of gripper has the possibility of picking items. However, bins are usually crowded. In a crowded bin, there are no gaps to insert fingers for picking. Therefore, grippers which have a possibility of collisions, such as multi-finger grippers, may be hard to use at the early stage. Vacuum or suction grippers are better to use than the finger-based grippers. In a crowded bin, target items may be under other items. Therefore, vacuum or suction grippers with a small-sized cup are more suitable to pick items at the early stage.

In general, the state inside the bins toward the end of the picking process changes as follows:

- Items become sparse and isolated from each other.
- Items hard to recognize and pick tend to remain.
- Items tend to remain near the walls of the bin.

When items are isolated, grippers which have a possibility of collision, like two-fingers, may reach a grasp pose more easily. Further, the more sparse the items get, the risk of picking several items also decreases, and approaching items in hard-to-pick poses becomes easier. Thus, large-sized suction (vacuum) grippers become easy to use.

In the state transition of bins as described above, the robot should switch a gripper combination when items are placed sparsely. We define the **Object sparseness** S as below, for understanding the states of bins.

$$S = S_B - N\pi R^2 \,, \tag{1}$$

where S_B denotes the area of the base of a bin, R denotes an average radius of circles which circumscribe items, and N denotes the number of items. N is calculated by subtracting the number of items taken out from the known total number of items in the bin.

We switch the gripper combination according to whether S has a positive value or not. If S is positive, gaps between items are visible inside the bin. If S is negative, items are piled in the bin with hard-to-use gaps.

4.2 Combination of multiple grippers

We propose a robot system which can properly use two gripper combinations. The proposed robot system is shown in Fig. 2. Two robot arms are mounted on linear sliders, facing each other with the item bins in the center in between them. Each robot arm is able to operate individually and has an RGB-D sensor and a force sensor. The RGB-D sensor is used for item and picking position detection.

The recognition algorithm is described below. The force sensor is used for force control when the robot picks and places items. The proposed robot system has three different types of gripper: suction, vacuum, and two-finger. The suction gripper is mounted on the left-side robot, as shown in Fig. 2. The vacuum and two-finger grippers are mounted on the right-side robot, as shown in Fig. 2. The two-finger gripper is used after removing the vacuum gripper by using a tool changer mechanism. These three different types of gripper were judged to be necessary to surely pick all 40 items given prior to the competition. In order to use three types of gripper for two robot arms, a gripper combination strategy based on the object sparseness is used. The three types of grippers on the proposed system are shown in Fig. 3.

Suction gripper. This gripper is based on a blower. It suctions large amounts of air and lifts up an item so that it sticks onto the suction pad. This gripper is able to stably grasp materials such as cloth through which air can pass. Although it cannot pick items such as a mesh cup which completely lets air pass through, it can pick items of almost any kind, shape, and material when compared to the other gripper types. To increase the versatility for picking items, the tip of the proposed suction gripper is covered with a large flexible pad, as shown in Fig. 3(a). As this pad wraps "along" the surface of items and creates a caging state, it is able to stably pick various items. One of its disadvantages is that the risk of picking several items simultaneously increases as the pad is soft and easy to deform. Moreover, the gripper also tends to pick other items nearby depending on the strength of the air flow. For this reason, recognizing other nearby items which may interfere in the picking is important.

Vacuum gripper. This gripper is based on a vacuum pump. It is highly reliable

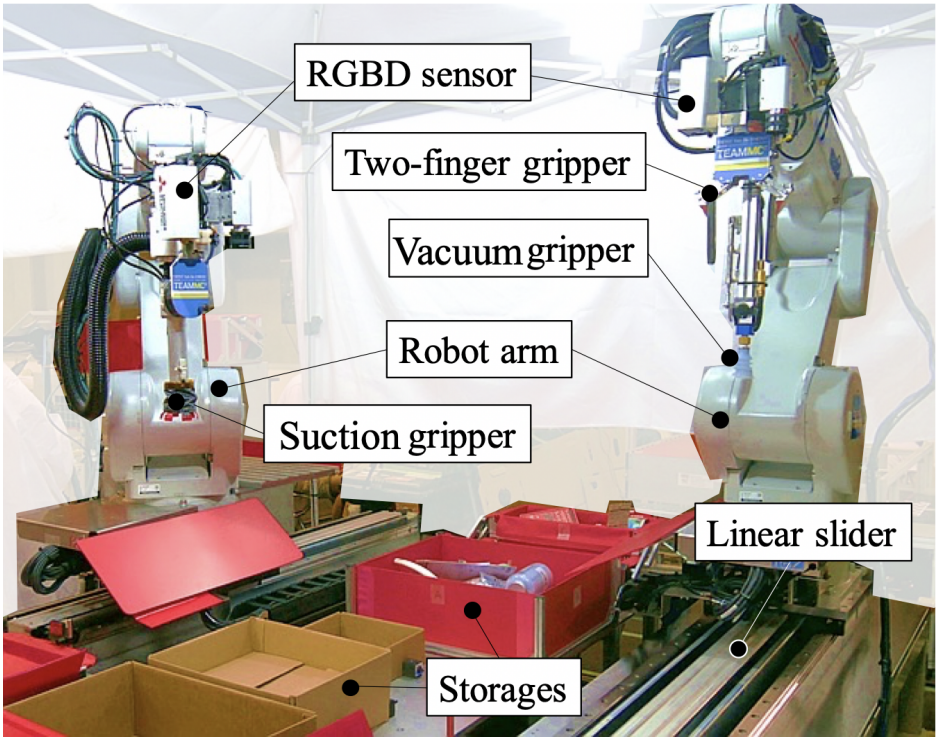

Figure 2: Proposed bin-picking robot system for various mixed items. The system has three different types of gripper: vacuum, suction and two-finger. SSD-based object detection and 3D-pose estimation algorithms can detect the graspable items and the grasping points. A multi-gripper switching strategy based on our proposed metric as named object sparseness is applied to the system. Thus, the proposed system can switch grippers according to the state of the bins.

and is widely used in factory automation. When the vacuum gripper comes into full contact with the surface of an item, it suctions the air between the closed pad and the item surface, which generates negative pressure. This negative pressure is higher than that of a suction gripper, which enables stable picks of the items even when using a small diameter pad. The vacuum gripper shown in Fig. 3(b) is designed with a pad with smaller diameter than that of the suction gripper shown in Fig. 3(a). As a result, it becomes easier to create a tight seal with the surface of small items and reduces the risk of picking several items simultaneously. Even so, it is unable to pick items when negative pressure cannot be generated on its surface. Examples are items for which the vacuum pad cannot completely come into full contact with their surface or items such

Figure 3: Three types of grippers on the proposed system: (a) suction gripper for various items, (b) vacuum gripper for items under other items, and (c) two-finger gripper for mesh items and hard-to-pick items.

as clothes which allow even a tiny amount of air to pass through. Therefore, this is a method suitable to pick items with pinpoint accuracy though the types of item will be limited.

Two-finger gripper. This gripper is able to pinch and pick items even if air passes through their surface. The fingers must open wide to pick various items. Therefore, the fingers of the proposed two-finger gripper, shown in Fig. 3(c) open wide by attaching a link mechanism to the tip of commercially available parallel chuck. Although the two-finger gripper is able to pick many kinds of items, there is a caveat. In contrast to suction and vacuum grippers which stably pick the items once they find the surface of the item, the position of the gripper must be determined while taking into consideration the collision with nearby items. For this reason, higher accuracy is required for grasp point determination methods based on RGB-D sensors.

Gripper combination. As we mentioned in Section 4.1, two-finger grippers have the possibility of collision. Therefore, we use a suction and a vacuum gripper at the early to middle stages. Then, we switch the gripper combination of a suction and a two-finger gripper by using the **Object sparseness** S.

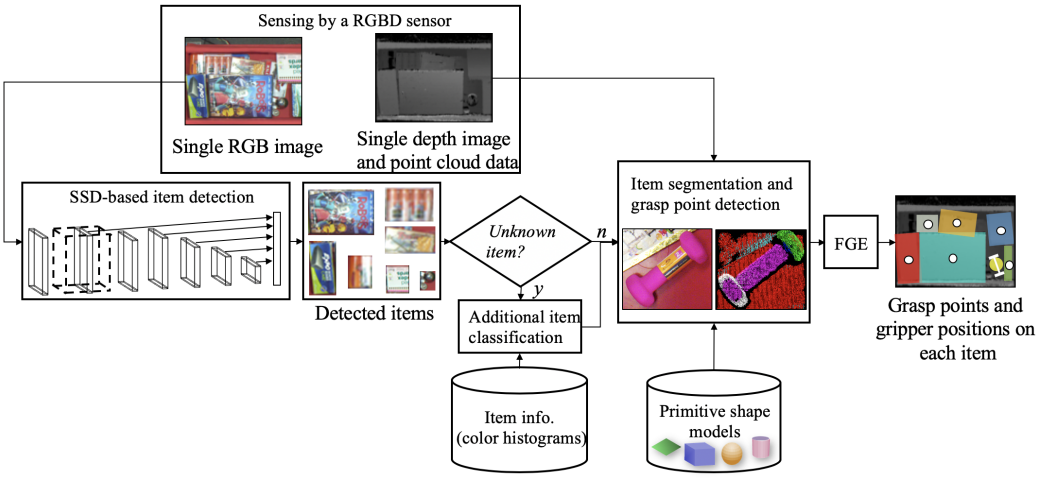

Figure 4: Proposed recognition flowchart for picking target items in bins.

4.3 Item recognition

We describe the item recognition method of the proposed robot system, which uses an RGB-D sensor, as shown in Fig. 4. There are three large phases in the proposed item recognition flowchart.

Item detection. Bounding boxes and multi-class recognition takes place simultaneously for several times on a Single Shot MultiBox Detector (SSD) based network [30]. As shown in Fig. 5, apart from the general index in SSD, box offset estimator and item classifier, "objectness" is adopted and evaluated at the same time. **Objectness** is an our proposed classifier which determines whether an item in a bounding box is an object or just background or a shadow. Bounding boxes classified as non-object with high confidence by this classifier are rejected and item classifications are given to the remaining boxes. By doing this, items are detected with high accuracy in a bin with mixed items. Also in SSD, items classified with low accuracy or unknown to the SSD network are further identified by color histogram obtained from the visible portion of the item. From the above, bounding boxes containing the items detected from RGB images are classified.

Item shape and pose estimation. Unknown items with no prior information are also the subject of bin picking. That is why pose estimation using 3D CAD models of items is not used. Instead, we assume the shape and pose of items, *i.e.*, segments, with a method [35] in which point cloud data and primitive shape models are matched.

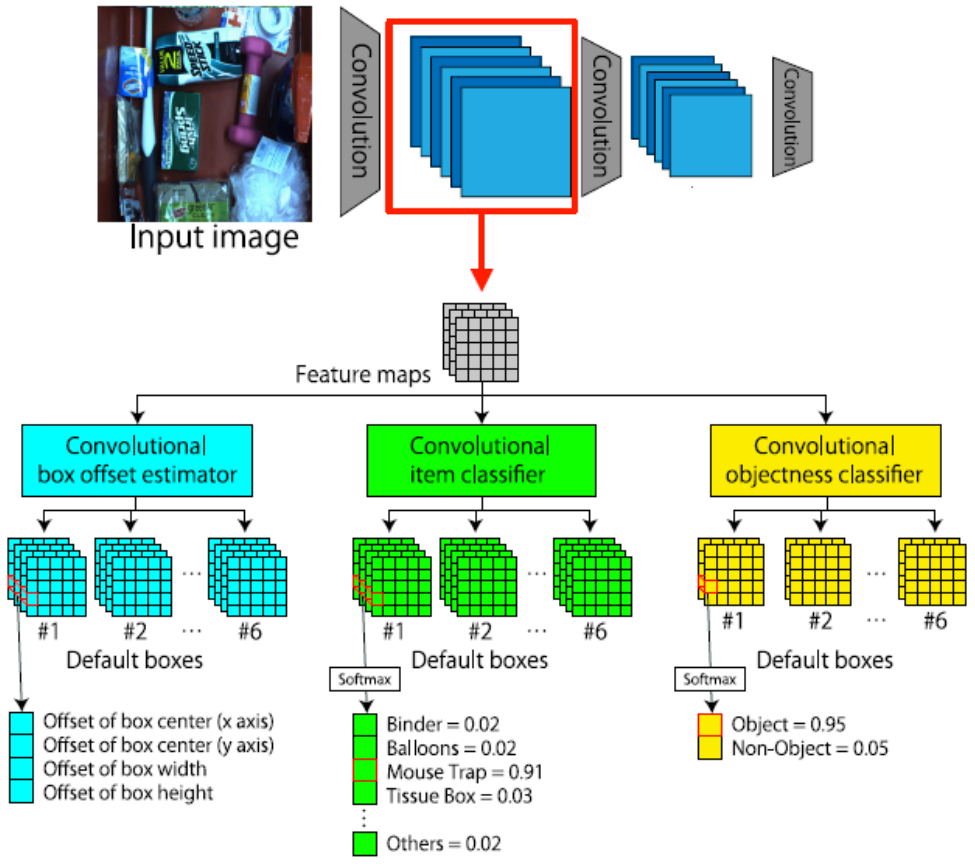

Figure 5: SSD-based item detection method. *Objectness* is our proposed classifier and is applied to the original SSD model to reduce the detection error caused by the background.

Grasp point detection. The point where the pad of suction and vacuum grippers is closer to the center of gravity of a fitted primitive shape is taken as the grasp point. We use Fast Graspability Estimation (FGE) [17] to determine the collision with nearby items for the two-finger gripper. This method enables to pick unknown items without shape model or even when the item is recognized as unknown as long as the gripper models are provided.

Items both already learned and not sufficiently learned (those given in the last minute) are recognized and grasp positions and poses for suction, vacuum, and two-finger grippers are calculated without any item model. In the end, classified segments

Figure 6: Actual recognition result examples which were captured using two RGB-D sensors on each robot arm. Top-left: initial state of a part of the bin. Bottom-right: final state of a part of the bin. The bounding boxes are the detected items. The small circles are candidates of graspable points. A grasp point or two-finger gripper position is determined after the recognition process.

and gripper poses on each segment are calculated. Fig. 6 shows examples of recognition results with classified bounding boxes, and where suction and vacuum grasp points are detected. The grasp point for the two-finger gripper is calculated by applying this result to FGE.

For each item, the gripper to use when picking is prioritized in advance considering the characteristics of the item and each gripper. After recognizing an item, the system basically selects the gripper according to the priority in the current gripper combination and tries to pick the item.

4.4 Performance evaluation using successful picking rate

As we mentioned in Chapter 2, we participated as team MC^2 in the Amazon Robotics Challenge 2017 with a robot system which is described in Section 4.2. Our robot system picked 18 out of 20 items in the aforementioned Stow task and came out third. The second place was team Nanyang [14] which used two robot arms with suction grippers. They picked 16 out of 20. The first place, team MIT-Princeton [13] picked all 20 out of 20 with a gripper system in which multiple picking motions are gathered into one. The organizer of the competition selected the items for each team to pick.

Although the selection of items varied among teams, the proportion of hard-to-pick items was considered to be similar.

We evaluated the performance of the proposed system as a bin-picking robot system. For comparison, we chose team MIT-Princeton's system [13] which is similar to ours (*i.e.*, they solved the problem as the hard bin-piking problem by using a multi-gripper strategy). We evaluated the successful picking rate from the competition videos. We analyzed the grasp rate of each gripper.

Successful picking rate. The successful picking rate shown in Fig. 7 is defined as follows: the success rate for all pickings when it is a success if an item is picked from the bin and placed without dropping; and it is a failure if an item is missed or several items are picked simultaneously. The transition of Team MIT-Princeton [13] compared to the proposed system is shown below. As shown in Fig. 7, the proposed system, which switches gripper combination based on the object sparseness, maintains a high successful picking rate from the beginning. Although the rate falls in the second half, it is consistently higher than the compared system.

The details of the number of picked items and the successful picking rate of each gripper are shown in Table 1. The final successful picking rate is 64.3% for the proposed system and 59.4% for the compared system. The vacuum gripper in the proposed system obtained 100% of successful picking rate. The two-finger gripper in the compared system also recorded a high successful picking rate of 75%. In both systems, the suction gripper succeeded to pick most items. Here, the number of picked items for the compared system is 19 because a point was given to an item dropped while it was transported as it coincidentally ended up in the right bin. In accordance with the definition of successful picking rate mentioned above, this is counted as a failure. Picking the item in Fig. 8(a) was tried for many times (from trial 16 to 25). In the end, the two-finger gripper succeeded to pick the hard-to-pick items, as shown in Fig. 8(a) and (c).

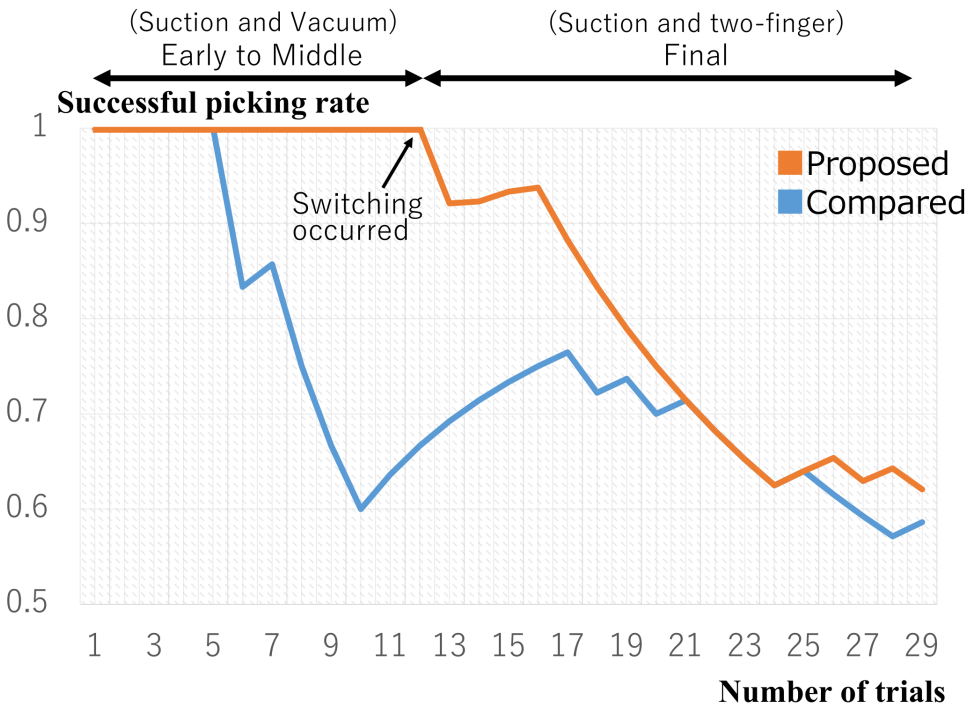

Figure 7: Transition of the successful picking rate. The orange line corresponds to the proposed system. The proposed system uses a suction gripper and a vacuum gripper from the beginning to the middle. After 12 picks, it switches to a suction gripper and a two-finger gripper based on the object sparseness.

Table 1: Number of picked items and the successful picking rate of each gripper of the two systems.

Proposed system				
	Suction	Vacuum	Two-finger	Total
Number of picked items	9	6	3	18
Successful picking rate [%]	64.3	100	37.5	64.3
Compared system				
	Suction	Vacuum	Two-finger	Total
Number of picked item	13	-	6	19
Successful picking rate [%]	54.2	-	75	59.4

(a) (b) (c)

Figure 8: Items which are hard to recognize and pick: (a) shiny metal parts, (b) translucent cup, and (c) black mesh cup. The proposed system could ultimately pick these items but several picking errors occurred.

Chapter 5

Performance metrics to evaluate bin-picking robot Systems

5.1 Proposed set of performance metrics

In Section 4.4, we have compared two systems (team MIT-Princeton and team MC^2 at the Amazon Robotics Challenge 2017) using successful picking rate as a performance metric. Mean Picks Per Hour (MPPH) [23] is currently widely-used as a bin-picking system performance metric [23, 37]. Since a high correlation to the score of the competition is seen with MPPH, it is suitable as a performance comparison metric of several systems developed for the Amazon Robotics Challenge 2017.

However, comparing system performance using such a single number means that only one aspect of the robot systems is considered and it is not clear which technologies and combinations are important. A solution to this problem is to evaluate the system performance in a comprehensive way by calculating several metrics for each system.

We propose to use **Number of trials per hour**, **Mean Time Between Failure (MTBF)**, **Mean Time to Repair (MTTR)**, and **Availability**, in addition to **Average probability of success** which is the same number as the successful picking rate and **MPPH** to evaluate the system performance. These introduced metrics, **MTBF**, **MTTR**, and **Availability** are originally used in a field of reliability engineering [5, 6].

By evaluating each system individually and comprehensively, we analyze their system design policy and system performance in a multifaceted manner. In each system, their subsystems work differently as they are designed differently, then multiple metrics will reflect the difference between systems.

MPPH is calculated by multiplying the **Number of trials per hour** and the **Average probability of success**. These two metrics try to analyze MPPH in more detail by breaking it down into two factors.

Average time per trial is measured from the competition video of each team. Then, **Number of trials per hour** can be calculated from the Average time per trial. These

two metrics can be used for comparison of a bare picking action speed.

The **Average probability of success** is also measured from the competition video of each team.

MTBF stands for Mean Time Between Failures and is calculated by dividing the normal operating time by the failure count of the system.

$$\mathrm{MTBF} = \frac{T_{\mathrm{up}}}{N_{\mathrm{down}}}, \tag{2}$$

where T_{up} is the duration of the system running well, and N_{down} is the number of times the system fails. These are obtained from the competition videos.

According to this definition, **MTBF** represents the average length of time that the system continues to work as expected. In other words, we can compare the length of time that the control algorithm is able to keep up with the variations of phenomena.

MTTR stands for Mean Time to Repair and is obtained as follows:

$$\mathrm{MTTR} = \frac{T_{\mathrm{down}}}{N_{\mathrm{down}}}, \tag{3}$$

where T_{down} is the duration from the beginning of the failures to the end of the recovery actions and is obtained from the competition videos.

This metric is expected to reflect the skills of each system to recover when a failure occurs. That is, we can compare the design strategies of the system as well as the **MTBF**.

Availability is obtained as follows:

$$\mathrm{Availability} = \frac{\mathrm{MTBF}}{\mathrm{MTBF} + \mathrm{MTTR}}. \tag{4}$$

Availability is a dimensionless quantity, which represents the ratio between the time that the system is running well and the total time the system is operating. We can expect a comparison of the duration of normal operation and recovery operation for each system. It is still a metric that is expected to reveal the design strategy of the system.

The purpose of introducing the set of performance metrics described above is primarily a comparison of the detailed functionality and performance of each system. The true goals are an uncovering of system design strategies, a discovery of measures to

obtain better functionality and performance, and a discovery of legitimate technologies for the target purpose.

5.2 Compared systems

5.2.1 Team MIT-Princeton (1st place)

The MIT-Princeton system [13] consists of a 6DOF ABB IRB 1600id robot arm next to four picking work-cells (see Fig. 9a). The robot arm uses a multi-functional gripper with two fingers (built on top of a Weiss WSG 50 gripper) for parallel-jaw grasps and a custom retractable suction cup. The gripper is designed to function in cluttered environments: finger and suction cup length are specifically chosen such that the bulk of the gripper body does not need to enter the cluttered space. One gripper fingertip is equipped with a GelSight tactile sensor, while the other fingertip uses an actuated fingernail for scooping along the sides of storage bins. Each work-cell consists of a storage bin, as well as four fixed-mounted RealSense SR300 RGB-D cameras: two cameras overlooking the storage bins (positioned on opposite sides) are used to infer grasp points, while the other two pointing towards the robot gripper (also positioned on opposite sides) are used to recognize objects in the gripper. Each work-cell also includes a force sensor underneath for 1) checking the weight of picked objects, and 2) detecting collisions.

The system is built around a grasp-first-then-recognize pipeline. For each pick-and-place operation, it uses fully convolutional networks (FCNs) to take as input RGB-D images of the work-cell, and output pixel-wise confidence scores (i.e., affordances) of four different motion primitives for picking (see Fig. 9b): top-down suction, side suction, top-down grasp, side-flush grasp. Each pixel of the output represents a suction or parallel-jaw grasp centered at the 3D location of that pixel ' s corresponding surface in view (Fig. 9c). The FCNs are trained using a dataset of 1,837 RGB-D images of cluttered work-cells, with good/bad grasp locations manually annotated by human experts. During inference, the system selects and executes the motion primitive with the highest predicted confidence score, picks up one object, isolates it from the clutter, holds it up in front of cameras, recognizes its category, and places it into the appropriate bin. The recognition algorithm uses a two-stream network to learn a common feature embedding space between 1) observed images of held objects, and 2) product images – where

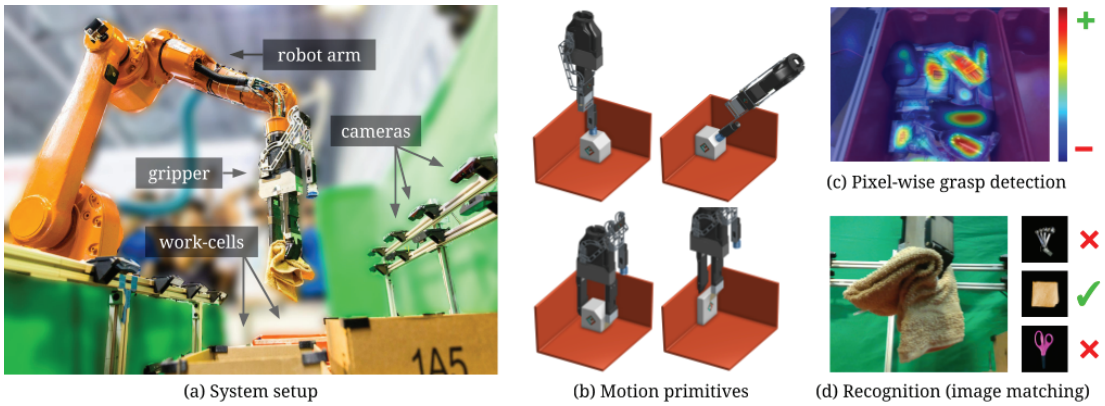

Figure 9: The MIT-Princeton system setup (a) consists of a 6DoF robot arm next to four picking work-cells. The system uses (c) FCNs to predict pixel-wise grasping confidences scores (*i.e.*, affordances) of (b) four motion primitives using suction and parallel-jaw grasping. After executing the motion primitive at the 3D location of the pixel with the highest confidence score, the system picks up an object and uses (d) a two-stream network to match images of the held object to the most similar product image for recognition.

images of the same object match to more similar output features. Since this network architecture does not rely on knowing the number of object categories beforehand, it is capable of recognizing images of novel objects unseen during training by matching them to corresponding product images that are provided at test time (Fig. 9d). Prior to the competition, the network is trained over observed-image-to-product-image pairs of known objects.

This system design has several advantages. First, the FCN-based grasping algorithm is model-free and agnostic to object identities. It detects grasps by using local geometric and texture features on objects, allowing it to learn biases that can generalize to novel objects without retraining (e.g. flat surfaces are good for suction, porous surfaces are bad for suction, etc.). Second, the object recognition algorithm works without task-specific data collection or retraining for novel objects, which makes it scalable for applications in warehouse automation and service robots where the range of observed object categories is large and dynamic. Third, the grasping framework supports multiple grasping modes with a multi-functional gripper (suction and grasping) and thus handles a wide variety of objects. Finally, the entire processing pipeline

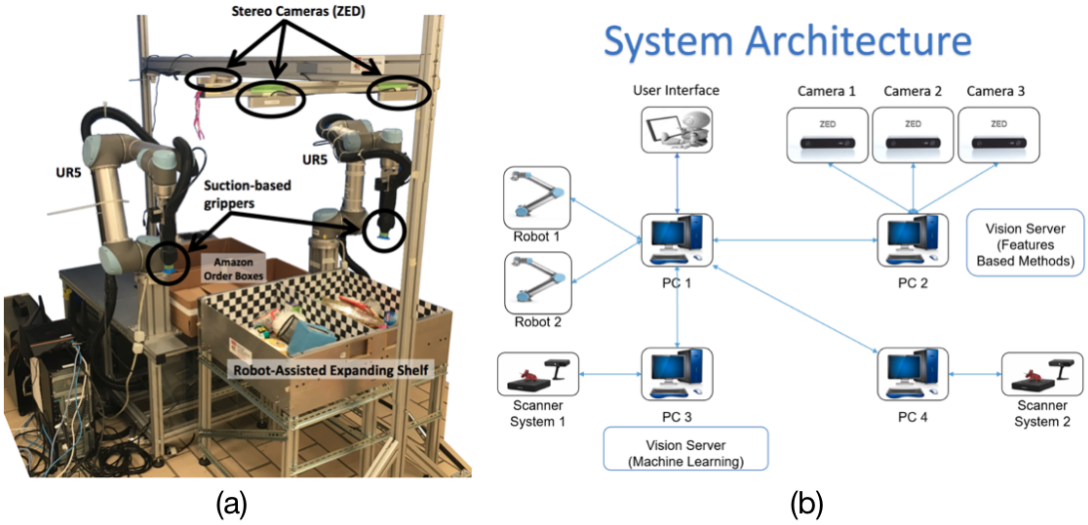

Figure 10: Team Nanyang's system (a) and its system architecture (b).

including grasp detection and recognition requires only a few forward passes through deep networks and thus executes quickly (a few hundred milliseconds in total per pick-and-place).

5.2.2 Team Nanyang (2nd place)

The team formed by members of the Nanyang Technological University (Singapore) developed a dual-arm robot system equipped with suction-based grippers and a top-open drawer-like storage system [14]. The robot system features two identical manipulators (Universal Robots UR5), three stereo cameras (Stereolabs ZED) and two custom-built grippers. The built system is shown in Fig. 10(a) together with its system architecture shown in Fig. 10(b).

The workspace is divided into two individual and one shared work cell to optimize the manipulation performance and decrease the risk of collision between the manipulators. The shelf has two bins which temporarily extend sideways in order to disperse the cluttered pile of items. This allows the system to have easier access to the items and to facilitate the object detection by decreasing occlusion.

For object detection, they use the results of either one of two classifiers, one based on engineered features and the other based on learned features, whichever has the

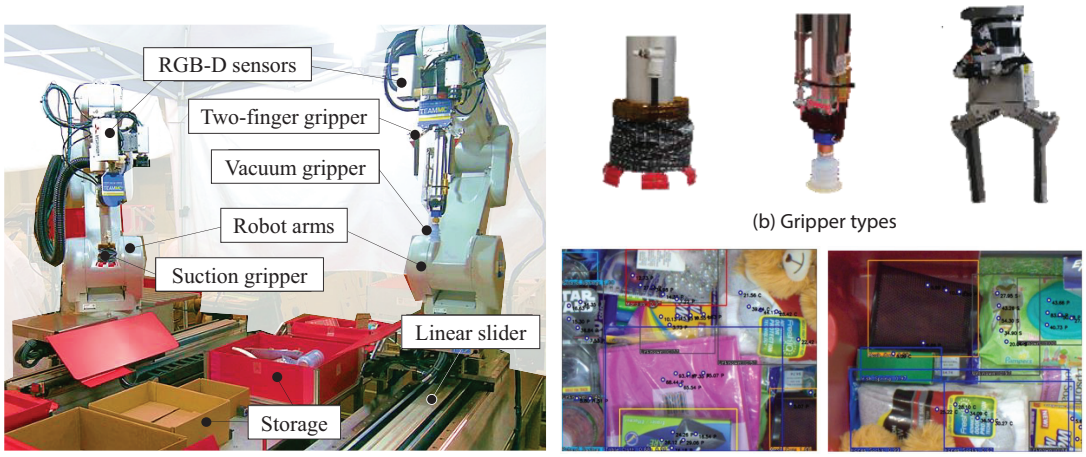

Figure 11: Robot system by team MC^2. (a) the system overview. (b) The system has three different types of gripper: suction, vacuum and two-finger. (c) SSD-based object detection and 3D pose estimation algorithms can detect the graspable items and the grasping points from cluttered scenes.

highest confidence. This is because they expect higher confidence for unknown items from the former, and higher confidence for known items from the latter. As engineered features, they use Grid-based Motion Statistics (GMS) [38], which is a feature detection algorithm similar in principle to SIFT but superior in performance. The learned-features are extracted using CNNs.

The grippers are suction-based since over 98% of the training items were successfully grasped using our modified suction cups. The grasping strategy consists mainly on approaching the objects straight down from the top, which is effective for almost 98% of the items.

5.2.3 Team MC^2 (3rd place)

The team MC^2 is explained again here for easy comparison with other teams, though it is described in section 4.2. The system is shown in Fig. 11(a). Two robot arms are mounted on linear sliders, facing each other with the item bins in the center in between them. Each robot arm is able to operate individually and has an RGB-D sensor and a force sensor. The RGB-D sensor is used for item and picking position detection. The

recognition algorithms are based on SSD, graspability, and primitive shape matching for item detection and classification, gripper pose detection, and item pose estimation separately as shown in Fig. 11(c). The force sensor is used for force control when the robot picks and places items. The proposed robot system has three different types of gripper: suction, vacuum and two-finger as shown in Fig. 11(b). The suction gripper is mounted on the left-side robot, as shown in Fig. 11(a). The vacuum and two-finger grippers are mounted on the right-side robot, as shown in Fig. 11(a). The two-finger gripper is used after removing the vacuum gripper by using a tool changer mechanism.

We devised a strategy in which the gripper combination changes accordingly. As bins are crowded and items are on top of each other, vacuum gripper, which picks items in smaller a surface area for picking, is preferred. As items are both large and small, the suction gripper, which is able to pick large items once it recognizes the surface, is also suitable. In contrast, collision due to item crowding inside the bin must be considered for the two-finger gripper and it is hard to obtain a pose for grasp positioning in a crowded bin. Therefore, the combination of vacuum and suction gripper is chosen for the beginning and middle stages of the picking task.

When items are isolated, two-finger gripper can reach a grasp pose more easily. Besides, the value of two-finger gripper rises because the remaining items are hard to pick with vacuum and suction grippers used at the early and middle stages. What is more, the more sparse the items get, the risk of picking several items also decreases, and approaching items in hard-to-pick poses becomes easier. Thus, the suction gripper is also adopted. To achieve the strategy explained so far, we configured a robot system in which one robot arm has a suction gripper and the other has vacuum and two-finger grippers, as described in [15]. The grippers are switched with a tool changer.

5.2.4 Team NAIST-Panasonic (4th place)

Team NAIST-Panasonic is formed by the Nara Institute of Science and Technology (NAIST) and Panasonic Corporation and include members with experience in robotics competitions [39].

The proposed solution consists of a 7-DOF robot arm (KUKA LBR iiwa 14 R820) with a custom-made end effector, a controlled space (recognition space) with four RGB-D cameras, and a shelf (storage system) with weight sensors underneath [40]. The setup of the proposed bin-picking solution is shown in Fig. 12a.

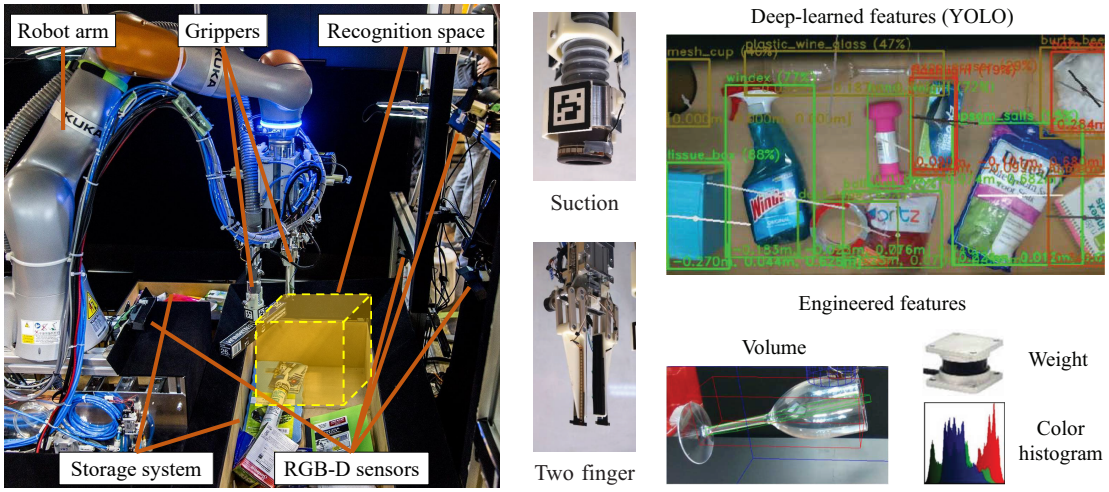

Figure 12: Bin-picking system proposed by team NAIST-Panasonic. (a) shows the system overview, (b) shows the suction and two-finger grippers, and (c) presents the learned and engineered features used for item recognition.

The end effector has a suction gripper and a two-finger gripper, shown in Fig. 12b, mounted on two separate linear actuators, and an RGB-D camera to recognize items and estimate the grasping points. The suction gripper consists of a compliant vacuum cleaner hose which is partially constrained to reduce swinging when transporting items. The two-finger gripper has high-friction rubber on its parallel fingers and is used as a secondary grasping tool. Both grippers include force-sensitive resistors to detect collisions with items and force control to avoid damaging items. The smart design of the end effector provided a reliable and consistent performance. The high flow and compliance of the suction gripper reduced the negative effects of vision and motion planning errors, making the system able to pick and transport the items safely.

The recognition space consists of four RGB-D cameras (Intel Realsense SR300) pointing at a space over the storage system, where eight LEDs control the illumination and the background of the cameras' views is controlled using non-reflective black plates. They combine learned and engineered features, shown in Fig. 12c, to achieve a robust object recognition for both known and unknown items. This was particularly useful in the case of the combination of bounding box volume and weight for clamshell-type and deformable items.

The strategy to recognize an item is: 1) point the end effector's camera to the target container, perform object detection using YOLO v2 [29] and grasping point estimation using RGB images, and pick the item with the highest recognition confidence; and 2) move this item into the recognition space to confirm or reject the initial belief using SVMs for single or combined engineered features (color histogram, bounding box volume, and weight) trained with data collected at approximately 90 seconds per item. A weight is assigned to each learned or engineered feature to adapt object recognition to the task requirements, physical characteristics of target items, and so on, resulting in a voting system that determines the final item class.

They designed the system to overcome failures by quickly detecting the most common errors and by preparing recovery behaviors in advance. This allowed them to retry failed grasping attempts in a short time. Moreover, they followed a time-saving strategy by 1) trying to pick the next viable item, if the previous attempt failed, without having to observe the scene again (*i.e.*, save time by reusing previous information from the end effector's camera), and 2) trying to pick an attempted but failed item by aiming to locations neighboring the original target grasping point. This strategy derives from the assumption that the scene has not changed significantly since the last attempt, and helps to compensate for vision sensing errors. Finally, the *recognize-while-holding* concept of the recognition space increased the robustness of the system to accidentally-dropped and unrecoverable items which could critically compromise the object recognition capabilities.

5.2.5 Comparison of system configurations

We show the system configurations of each system in Table 2. The main similarities which can be understood from this table are:

- All systems are based on industrial robots because accuracy and speed are important factors to complete the task. Industrial robot's high accuracy may be excessive but some collaborative robots are difficult to use for the Stow task because of low accuracy.

- Almost all teams based their grippers on suction (or vacuum) and two fingers. Suction-based grippers can pick many items including deformable objects but they are difficult to apply to mesh items (*i.e.*, air-permeable items). On the other

Table 2: System configuration of each team.

	MIT-Princeton	Nanyang	MC^2	NAIST-Panasonic
Robots	One 6-DoF robot arm	Two 6-DoF robot arms	Two 6-DoF robot arms on 1-DoF linear sliders	One 7-DoF robot arm
Sensors	Sixteen fixed RGB-D sensors, four force sensors (below bins), one tactile sensor (GelSight on gripper), and one air pressure sensor	Three fixed RGB-D sensors	Two RGB-D sensors and two force sensors on robot arms, and two weight sensors	Five RGB-D sensors (four fixed and one on robot arm), two weight sensors, two FSR-based contact sensors, and one air pressure sensor
Grippers	Multi-functional gripper with two fingers for parallel-jaw grasps and a retractable suction cup	Two suction-based grippers	One large suction gripper, one small vacuum gripper, and one two-finger gripper	One suction gripper and one two-finger gripper
Recognition algorithm	Two FCNs to infer grasping points for both suction and parallel-jaw grasping, and a two-stream network to match real images of objects to product images for classification	Mixed-mode classifier using feature extraction (GMS) and CNN	SSD-based item detector and classifier from a RGB image, gripper pose detector from a single depth map, and 3D pose estimator from a point cloud data	Multi-modal weighted voting classifier using learned and engineered features (YOLO from RGB, volume from depth, weight, and color histogram)
Unique features	Learning visual affordances for multi-functional gripping (grasping and suction)	Top-open extendable shelf design	Using three types of grippers and its combination strategy	High-flow suction gripper and fast failure recovery
# of robot arms	1	2	2	1
# of sensors	22	3	6	10
# of grippers	2	2	3	2

hand, two-finger grippers can pick mesh items. Therefore, the combination performs well.

- The item recognition of the systems is mainly based on RGB-D sensors and CNN-based algorithms. Open source computer vision implementations are easy to use for researchers of the robotics field and perform well enough.

The main differences are:

- The number of robots and their degrees of freedom are different. All teams basically pick items from above the storage system with 4 DoF. Therefore, 6 DoF should be enough. Robot systems can have many robots but systems with too many robots are hard to implement and are more prone to collision problems. Therefore, some teams use systems with fewer robots.

- Some teams used force sensors, weight sensors, visual-tactile sensors (GelSight), and so on. These sensors seem to be useful for the Stow task but the implementation may be difficult, mainly because there are very few useful open source projects to help with the implementation.

5.3 Comparison results using the set of performance metrics

We calculated the proposed set of performance metrics and relative values, as shown in Table 3. Here, Number of trials, Number of successes, Number of failures, Sum of up time, Sum of down time and total time were obtained from the videos recorded during the Amazon Robotics Challenge 2017. With regard to Sum of down time, the timing of returning to the normal flow after failing to grasp the object is confirmed. The timing of normal flow depends on the team. For example, in the case of the team MC^2, the timing is when the robot arm stopped for object recognition, and in the case of the team NAIST-Panasonic, the timing is when the approach was started to pick the next object.

Then, we normalized the most significant metrics based on the winning team (team MIT-Princeton), as shown in Table 4. We also show these normalized results in Fig. 13.

From Fig. 13, we observe that the score based on ARC rules, MPPH, and Availability are highly correlated, which makes it suitable as a comprehensive performance

Table 3: Results of metrics calculation of each team.

	MIT-Princeton	Nanyang	MC^2	NAIST-Panasonic
Score based on ARC* rules	**160**	125	120	110
Number of trials	32	37	28	**50**
Number of successes**	**19**	16	18	17
Number of failures	13	21	**10**	33
Successful picking rate	0.594	0.432	**0.643**	0.340
Average time per trial [sec]	23.1	24.3	32.1	**18.0**
Number of trials per hour	156	148	112	**200**
Average probability of success	0.594	0.432	**0.643**	0.340
MPPH	**92.6**	64.0	72.0	68.0
Sum of up time [sec]	**535**	504	488	437
Sum of down time [sec]	**204**	396	412	463
Total time [sec]	**739**	900	900	900
MTBF [sec]	41.2	24.0	**48.8**	13.2
MTTR [sec]	15.7	18.9	41.2	**14.0**
Availability	**0.724**	0.560	0.542	0.486

* Amazon Robotics Challenge 2017.

** Successful sequences of pick, move, and place.

Table 4: Normalized results of selected metrics based on highest score team (team MIT-Princeton).

	MIT-Princeton	Nanyang	MC^2	NAIST-Panasonic
Score based on ARC* rules	**1.00**	0.78	0.75	0.69
Average time per pick	1.00	1.05	1.39	**0.78**
Number of trials per hour	1.00	0.95	0.72	**1.28**
Average probability of success	1.00	0.73	**1.08**	0.57
MPPH	**1.00**	0.69	0.78	0.73
MTBF	1.00	0.58	**1.19**	0.32
MTTR	1.00	1.20	2.63	**0.89**
Availability	**1.00**	0.77	0.75	0.67

* Amazon Robotics Challenge 2017.

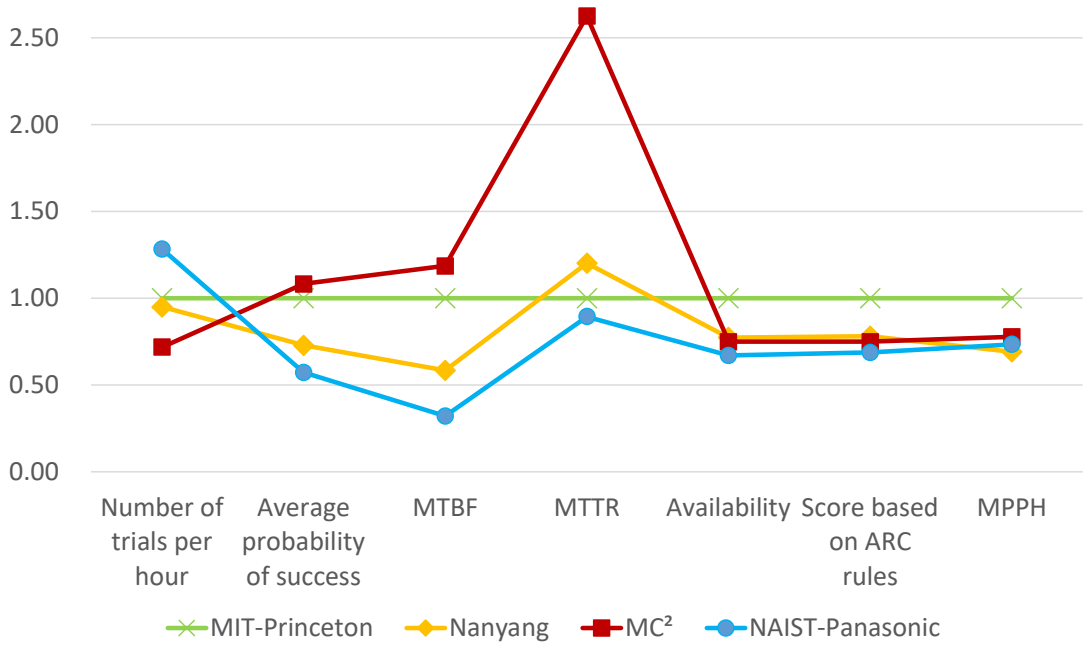

Figure 13: System performance comparison based on the selected metrics.

metric in that sense. We consider that the slight deviation is caused by a difference in scoring when including a bonus point. However, the other metrics are considerably fluctuating.

Fig. 14 shows the correlation of the selected metrics. The size of the circle, as well as the color difference, indicates the strength of the correlation. Availability and the score based on ARC rules have a particularly strong positive correlation. Average probability of success and MTBF also have a strong positive correlation. On the other hand, Number of trials per hour have a fairly strong negative correlation with Average probability of success, MTBF, and MTTR. The correlation is not strong except for the above.

Table 5 shows the details of the number of picked items and the successful picking rate of each gripper. Some parts of the table are described in section 4.4. Both Nanyang and NAIST-Panasonic have low successful picking rates.

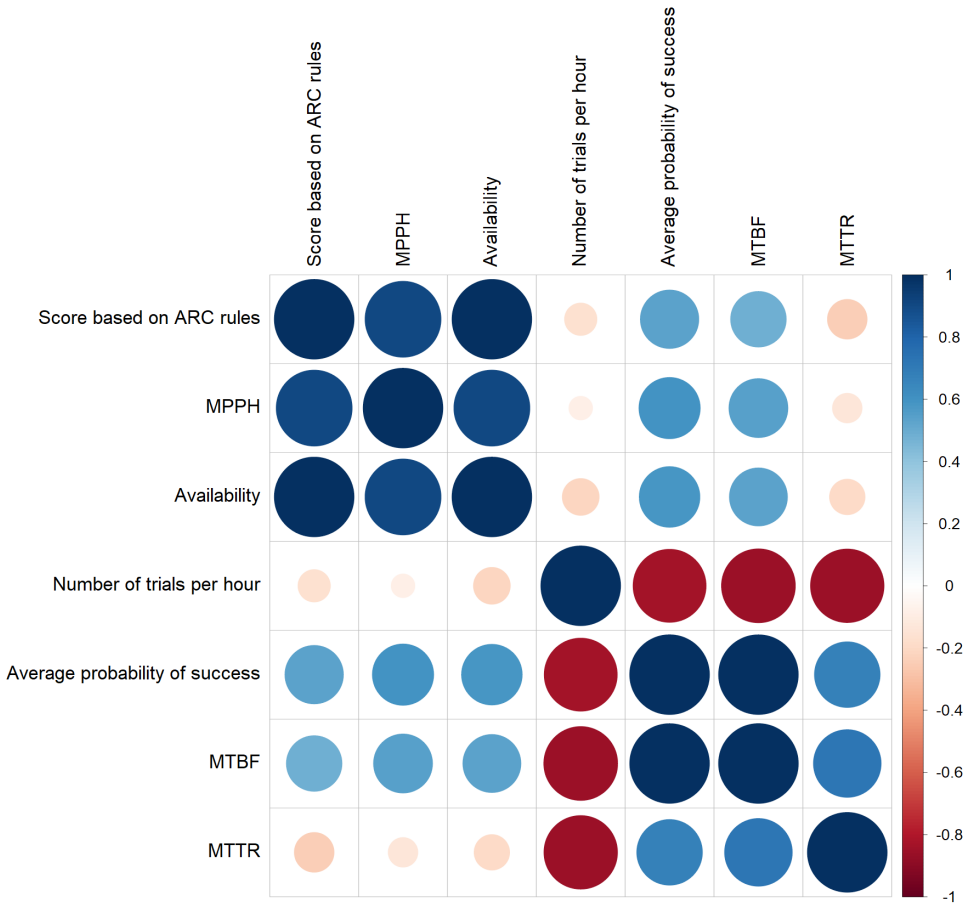

Figure 14: Correlation among the selected metrics.

Table 5: Number of picked items and the successful picking rate of each gripper. We refer to the blower-based suction as *suction* and vacuum-pump-based suction as *vacuum*. In this dissertation, dropped items during pick-and-place do not count as successful picking.

MIT-Princeton				
	Suction	Vacuum	Two-finger	Total
Number of picked items	13	-	6	19
Successful picking rate [%]	54.2	-	75	59.4

Nanyang				
	Suction	Vacuum	Two-finger	Total
Number of picked items	16	-	-	16
Successful picking rate [%]	43.2	-	-	43.2

MC^2				
	Suction	Vacuum	Two-finger	Total
Number of picked items	9	6	3	18
Successful picking rate [%]	64.3	100	37.5	64.3

NAIST-Panasonic				
	Suction	Vacuum	Two-finger	Total
Number of picked items	17	-	0	17
Successful picking rate [%]	34	-	0	34.0

Chapter 6

Discussion

6.1 Multi-gripper switching strategy

As shown in Fig. 7, the proposed system maintained a higher successful picking rate than the compared system, especially from the beginning to the middle of the picking task. As shown in Table 1, the vacuum gripper in the proposed system recorded 100% successful picking rate. The analysis showed that the vacuum gripper accurately reached small available surfaces and managed to pick items on top of others or items underneath other items deep in the bottom of the bin. With the suction gripper used alongside, the system correctly decided to use the suction gripper for large items and the vacuum gripper for small items or items with a small visible area, which, in consequence, led to a high success rate. From these results, we observed that using two sizes, a small vacuum cup and a suction cup worked reasonably well.

On the other hand, the successful picking rate for the two-finger gripper was as low as 37.5%. This is because the two-finger gripper was used toward the end of the task, when the only remaining items were hard to pick. A shiny metal object shown in Fig. 8(a) is the item which was hard to pick. There were errors in item recognition and grasp point detection because the depth map measurement for this item was not stable. The measurement quality of depth map is crucial for grasp point detection. Due to this measurement quality, picking the item in Fig. 8(a) was tried for many times (from trial 16 to 25). In the end, the two-finger gripper succeeded to pick the hard-to-pick items shown in Fig. 8. We can conclude that the strategy to pick the hard-to-pick items toward the end of the task worked.

From this point of view, our proposed switching strategy of a gripper combination worked reasonably well. In other words, the timing of switching the gripper combination was appropriately controlled based on the sign of the **Object sparseness** S. This time, we proposed the strategy to switch the gripper combination based on the sign of S. That is, the threshold value is set to zero, but if the threshold value is set to a value other than zero, the performance may change, and verification of this is an issue for the

future. We consider that the value of using two-finger and multi-finger grippers will increase in the future when the ability of item recognition improves and there are more varieties in grasp motion control.

It is worth mentioning that the successful picking rate of the two-finger gripper is high in the compared MIT-Princeton system [13], namely 75%. With a gripper which has a fingertip using an actuated fingernail for scooping along the sides of the bins as described in subsection 5.2.1, an item extremely hard to pick such as a book leaning on the wall of the bin was successfully picked. It obtained good results by specially devising the structure of the gripper.

6.2 Analysis of the performance comparison

In this section, we analyze the results shown in Fig. 13. We consider changes of each metric in comparison to the actual system implementation, and explore the system design concept.

MPPH is a good metric that represents system performance, as evident in the fact that **MPPH** and the score based on the rules of the Amazon Robotics Challenge are similar. Hereafter, we examine the factors constituting the **MPPH**, namely, the **Number of trials per hour** and **Average probability of success**.

First, **Number of trials per hour** is very low for MC^2, while it is significantly high for team NAIST-Panasonic. The other two teams are in the middle. When we look closely at the system design of each team, MC^2, for example, has a hand-eye system with a vision sensor attached to a wrist of their robot, and it is configured to perform the vision sensing operation and the other operation sequentially. In other words, the recognition operation is performed after the completion of the stow operation, which is one cycle before, then, the picking operation starts. Therefore, one cycle takes an amount of time while teams MIT-Princeton and Nanyang can perform the previous stow operation and the recognition operation concurrently and shorten their cycle times. Team NAIST-Panasonic also has a hand-eye system to recognize items and estimate the grasp points. Team NAIST-Panasonic has a strategy to shorten cycle times and increase the number of trials by detecting failures quickly and taking recovery actions immediately. The system reuses previous information instead of recognizing objects again in the event of a failure. From the beginning to the middle, this strategy worked well, but in the final stage it sometimes did not. These differences in

system design for each team are reflected in the number of trials per hour.

In terms of the **Average probability of success**, team MC^2 and team MIT-Princeton are comparable, while NAIST-Panasonic and Nanyang have lower values. As a system design concept, the difference lies in whether it is thought that every single operation is important or not. Team MC^2 has a strategy to switch the gripper combination of three different types of gripper appropriately to surely pick items and achieved the highest average probability of success. Team MIT-Princeton has a specially devised multi-functional gripper and four types of gripping motion primitive to increase the probability of success. On the other hand, team NAIST-Panasonic attempts to score by retrying quickly if it fails, assuming it will fail. There is a clear difference in the system design concept.

With regard to the **MTBF**, team MC^2 has the largest value, followed by team MIT-Princeton. Team NAIST-Panasonic has a particularly small value. This metric shows how long the robot system can continue to operate normally on average before failing. Since MTBF has a strong correlation with the average probability of success, as shown in Fig. 14, MTBF also shows the system design concept of each team. Team MC^2 and team MIT-Princeton try to keep normal operation as long as possible, while the other two teams allow failures.

Considering **MTTR**, team MC^2 has a particularly large value compared to the other three teams. Note that the smaller the value, the better. This metric shows how quickly the robot system can return to normal operation if failures occur. As team MC^2 has a hand-eye system, performs the vision sensing operation and the other operations sequentially, and has no quick recovery strategy, it takes quite a long time to return to normal operation once a failure occurs. It can be said that the system design of team MC^2 is very disadvantageous from the viewpoint of not only short operation time but also quick recovery from failure. NAIST-Panasonic has the smallest value. Their design policy of detecting failures quickly and retrying in a short time is reflected in this value.

In terms of **Availability**, team MIT-Princeton has the highest value, and the other teams have similar ones. While the other teams are better in the metrics mentioned above, team MIT-Princeton seems to be balanced on the speed of operation, the probability of success, and the time of recovery from failure. Although team MC^2 has a high average probability of success, it took a long time to retry a failed attempt, that is, it had

a large MTTR and performed only gentle and naive actions, which resulted in repeated unsuccessful picking motions and low availability. Team Nanyang and team NAIST-Panasonic have lower MTTR than team MC^2, but more failures and lower average probability of success resulted in lower availability similar to team MC^2. Availability has a strong correlation with the score based on ARC rules. So, it can be said that the ARC rules evaluate availability of the system. Since Availability indicates the ratio of the time during normal operation to the total operating time, it is reasonable to evaluate it in the competition from a practical point of view. MPPH is also correlated with the score based on the ARC rules. In this sense, Availability and MPPH are suitable as comprehensive performance metrics for bin-picking robots in the fields of warehouse automation.

6.3 Lessons learned and important technologies for practical use

In the previous section, we estimated the design policy differences and their performance with newly introduced metrics, which could not be understood from the analysis based on a single metric.

We found that it is important to realize 1) short operation time, 2) high probability of success, and 3) short recovery time from failures. Achieving these three points at a high level will result in high availability and high MPPH, and the resulting robot system will be close to practical use. Now we discuss the technologies that are important for achieving these goals.

In terms of short operation time, the facts indicate that a hand-eye system is disadvantageous for shortening operation time because measurement and real work must be carried out in series, and parallelization is not possible. In other words, vision sensors should be placed separately to make measurement and actual work proceed simultaneously. When a hand-eye configuration is adopted, the vision sensors can be moved within the reach of the robot, and it is advantageous that the visual field can be moved and widened. On the other hand, in the separation type, the vision sensors have fixed arrangements, and the visual field can not be freely changed. If the robot system continues to work in the same environment, it will not be a problem. However, it could be disadvantageous if a hardware reconfiguration is needed.

The operation time of the system can be also shortened by using multiple robot arms in an appropriate configuration. If multiple robot arms can work simultaneously

in the same operation area while avoiding collisions, the operation time will be even shorter. In addition, shortening the robot trajectories by using robot arms with redundant degrees of freedom or making the robot motion itself faster will also lead to shorter operation time. At this time, we must be careful about item damage. In the competition, damage to an item results in a large deduction in score. Because there is no record of damage to items, it is not discussed in this dissertation. It is presumed that all teams are afraid of penalty points so that they added a wide-enough margin to their system operation. In a practical system, both elimination of item damage and short operation time must be achieved.

As for high probability success, it is of course important to improve recognition performance of vision systems, but it is also necessary to advance grasping technologies. In order to grasp various items, a gripper suitable for each item should be used. Since items can take various postures, the gripper also needs various approach motions. From the viewpoint of a practical use, it is very important to grasp an item without damaging it. In some cases, two robot arms may need to work together. For example, one robot may slightly move an item and the other robot may retrieve an item below it.

With regard to short recovery time from failures, how quickly to detect a failure and take the next action is important. In addition, from the perspective of increasing availability, recovery strategies for grasping failures are important in order not to repeat the same failure. Retrying which does not rely on probabilistic phenomena is required, for example, by changing parameters of the recognition algorithm and grasping method, and possibly changing postures of the object by such as flipping them, shaking the bins, and so on.

As is common to all of the above, cost and reliability are most important from a practical point of view. As the system becomes more complex, it becomes more costly, less reliable, and less practical. This must always be kept in mind in order to realize practical bin-picking robots for warehouse automation.

Chapter 7

Conclusions

In this dissertation, important technologies to realize practical bin-picking robots for warehouse automation are discussed.

First, a strategy to change the gripper combination during a bin-picking task based on the sparseness of objects inside bins, and a bin-picking robot system using it are proposed. The evaluation results using successful picking rate as a metric are shown, and the effectiveness of this strategy is verified.

Then, a novel set of performance metrics to evaluate bin-picking robots for various mixed items from multiple aspects is proposed. We quantitatively analyzed four robot systems developed for the Amazon Robotics Challenge 2017 using the proposed set and clarified hidden features and system design concepts behind the competition scoring. We showed the difference between these systems and clarified the importance of short operation time, high probability of success, and short recovery time from failures. The results may be qualitatively obvious, but for the first time they are quantitatively confirmed for bin-picking robots for various mixed items by the proposed set of performance metrics.

Finally, important technologies to realize practical bin-picking robots are discussed. Further technology developments are needed. We expect this analysis to be a good reference for advanced future technologies along with novel needs of the industry.

Acknowledgements

This dissertation could not have been completed without the guidance and assistance of many people.

First, I wish to express my sincerest gratitude to Professor Tsukasa Ogasawara. He strongly recommended me to obtain a Ph.D. degree and kindly instructed me. When I was about to give up my degree due to the difficulty of balancing work and research, he gently encouraged me and led me to this far goal.

I am most grateful to Assistant Professor Gustavo Alfonso Garcia Ricardez. He was very kind and courteous in helping me with my dissertation over and over again. He also gave me great advice many times based on his extensive knowledge and experience.

I would like to express my deepest gratitude to Professor Akio Noda of the Osaka Institute of Technology. His insightful advice has made great progress in this research. He always encouraged me when faced with challenges.

I would also like to express my deepest gratitude to Dr. Yukiyasu Domae of the National Institute of Advanced Industrial Science and Technology. When he belonged to Mitsubishi Electric Corporation, he led the development of our robot system as the leader of Team MC^2, and succeeded in 3rd place in the Stow task of the Amazon Robotics Challenge 2017, giving me the opportunity to write my dissertation. He also gave me a number of very helpful suggestions and advice for my dissertation.

I wish to deeply thank to my thesis committee: Professor Tsukasa Ogasawara, Professor Kenji Sugimoto, Associate Professor Jun Takamatsu, and Assistant Professor Gustavo Alfonso Garcia Ricardez for their helpful comments and suggestions.

I would like to express my sincere gratitude to Professor Hironobu Fujiyoshi of Chubu University, Professor Manabu Hashimoto of Chukyo University, Dr. Rintaro Haraguchi and Mr. Koji Shiratsuchi of Mitsubishi Electric Corporation, and many other members who worked together as members of Team MC^2.

Lastly, I would like to thank my loved family from the bottom of my heart for their immense support and understanding. They understood me trying to get a Ph.D. degree after I was over 60 years old, and they were very supportive. My wife, Reiko, in particular, staying at home due to COVID-19, always watched me gently, warmly and half-amazedly.

References

[1] Jeremy A. Marvel, Kamel Saidi, Roger Eastman, Tsai Hong, Geraldine Cheok, and Elena Messina, "Technology readiness levels for randomized bin picking," in *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*. New York, NY, USA: ACM, 2012, pp. 109–113.

[2] NIST, "Agile Robotics for Industrial Automation Competition," 2019. [Online]. Available: https://www.nist.gov/el/intelligent-systems-division-73500/agile-robotics-industrial-automation-competition

[3] Yasuyoshi Yokokohji, Yoshihiro Kawai, Mizuho Shibata, Yasumichi Aiyama, Shinya Kotosaka, Wataru Uemura, Akio Noda, Hiroki Dobashi, Takeshi Sakaguchi, and Kazuhito Yokoi, "Assembly challenge: a robot competition of the Industrial Robotics category, World Robot Summit - summary of the pre-competition in 2018," *Advanced Robotics*, vol. 33, no. 17, 2019, pp. 876–899.

[4] Mizuho Shibata, Hiroki Dobashi, Wataru Uemura, Shinya Kotosaka, Yasumichi Aiyama, Takeshi Sakaguchi, Yoshihiro Kawai, Akio Noda, Kazuhito Yokoi, and Yasuyoshi Yokokohji, "Task-board task for assembling a belt drive unit," *Advanced Robotics*, vol. 34, no. 7-8, 2020, pp. 454–476.

[5] "Reliability and availability basics," 2019. [Online]. Available: http://www.eventhelix.com/RealtimeMantra/FaultHandling/reliability_availability_basics.htm

[6] "Reliability engineering," 2019. [Online]. Available: https://en.wikipedia.org/wiki/Reliability_engineering

[7] Nikolaus Correll, Kostas E Bekris, Dmitry Berenson, Oliver Brock, Albert Causo, Kris Hauser, Kei Okada, Alberto Rodriguez, Joseph M Romano, and Peter R Wurman, "Analysis and observations from the first Amazon Picking Challenge," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, 2018, pp. 172–188.

[8] Carlos Hernandez, Mukunda Bharatheesha, Wilson Ko, Hans Gaiser, Jethro Tan, Kanter van Deurzen, Maarten de Vries, Bas Van Mil, Jeff van Egmond, Ruben Burger *et al.*, "Team Delft's robot winner of the Amazon Picking Challenge 2016," in *Robot World Cup*. Springer International Publishing, 2016, pp. 613–624.

[9] Yukiyasu Domae, Ryosuke Kawanishi, Koji Shiratsuchi, Rintaro Haraguchi, Masahiro Fujita, Yuji Yamauchi, Takayoshi Yamashita, Hironobu Fujiyoshi, Shuichi Akizuki, and Manabu Hashimoto, "Picking robot based on a two-finger gripper for various mixed items in shelves," *Journal of the Robotics Society of Japan*, vol. 38, no. 1, 2020, pp. 95–103, (in Japanese).

[10] Hironobu Fujiyoshi, Takayoshi Yamashita, Shuichi Akizuki, Manabu Hashimoto, Yukiyasu Domae, Ryosuke Kawanishi, Masahiro Fujita, Ryo Kojima, and Koji Shiratsuchi, "Team C2M: Two cooperative robots for picking and stowing in Amazon Picking Challenge 2016," in *Advances on Robotic Item Picking: Applications in Warehousing & E-Commerce Fulfillment*, Albert Causo, Joseph Durham, Kris Hauser, Kei Okada, and Alberto Rodriguez, Eds. Springer International Publishing, 2020, pp. 101–112.

[11] Douglas Morrison, Adam W. Tow, M. McTaggart, R. Smith, N. Kelly-Boxall, S. Wade-McCue, J. Erskine, R. Grinover, A. Gurman, T. Hunn *et al.*, "Cartman: The low-cost Cartesian manipulator that won the Amazon Robotics Challenge," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7757–7764.

[12] Gustavo Alfonso Garcia Ricardez, Lotfi El Hafi, and Felix von Drigalski, "Standing on giant's shoulders: Newcomer's experience from the Amazon Robotics Challenge 2017," in *Advances on Robotic Item Picking*, Albert Causo, Joseph Durham, Kris Hauser, Kei Okada, and Alberto Rodriguez, Eds. Springer International Publishing, 2020, pp. 87–100.

[13] Andy Zeng, Shuran Song, Kuan-Ting Yu, Elliott Donlon, Francois R. Hogan, Maria Bauza, Daolin Ma, Orion Taylor, Melody Liu, Eudald Romo *et al.*, "Robotic pick-and-place of novel objects in clutter with multi-affordance grasp-

ing and cross-domain image matching," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.

[14] Albert Causo, Zheng-Hao Chong, Ramamoorthy Luxman, Yuan Yik Kok, Zhao Yi, Wee-Ching Pang, Ren Meixuan, Yee Seng Teoh, Wu Jing, Hendra Suratno Tju *et al.*, "A robust robot design for item picking," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7421–7426.

[15] Masahiro Fujita, Yukiyasu Domae, Ryosuke Kawanishi, Kenta Kato, Koji Shiratsuchi, Rintaro Haraguchi, Ryosuke Araki, Hironobu Fujiyoshi, Shuichi Akizuki, Manabu Hashimoto, Gustavo Alfonso Garcia Ricardez, Albert Causo, Akio Noda, , Haruhisa Okuda, and Tsukasa Ogasawara, "Bin-picking robot using a multi-gripper switching strategy based on object sparseness," in *2019 IEEE International Conference on Automation Science and Engineering*. IEEE, 2019, pp. 1540–1547.

[16] Amend John, Brown Eric, Nicholas Rodenberg, Heinrich Jaeger, and Hod Lipson, "A positive pressure universal gripper based on the jamming of granular material," *IEEE Transactions on Robotics*, vol. 28, no. 2, 2012, pp. 341–350.

[17] Yukiyasu Domae, Haruhisa Okuda, Yuichi Taguchi, Kazuhiko Sumi, and Takashi Hirai, "Fast graspability evaluation on single depth maps for bin picking with general grippers," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1997–2004.

[18] Yun Jiang, Stephen Moseson, and Ashutosh Saxena, "Efficient grasping from RGBD images: Learning using a new rectangle representation," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 3304–3311.

[19] Ian Lenz, Honglak Lee, and Ashutosh Saxena, "Deep learning for detecting robotic grasps," *International Journal of Robotics Research*, vol. 34, no. 4-5, 2015, pp. 705–724.

[20] Lerrel Pinto and Abhinav Gupta, "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 3406–3413.

[21] Sergey Levine, Peter Pastor, Alex Krizhevsky, Julian Ibarz, and Deirdre Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *International Journal of Robotics Research*, vol. 37, no. 4-5, 2018, pp. 421–436.

[22] Alberto Rodriguez, Matthew T Mason, and Steve Ferry, "From caging to grasping," *International Journal of Robotics Research*, vol. 31, no. 7, 2012, pp. 886–900.

[23] Jeffrey Mahler, Florian T Pokorny, Brian Hou, Melrose Roderick, Michael Laskey, Mathieu Aubry, Kai Kohlhoff, Torsten Kröger, James Kuffner, and Ken Goldberg, "Dex-Net 1.0: A cloud-based network of 3D objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1957–1964.

[24] Jeffrey Mahler, Jacky Liang, Sherdil Niyaz, Michael Laskey, Richard Doan, Xinyu Liu, Juan Aparicio Ojea, and Ken Goldberg, "Dex-Net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," 2017. [Online]. Available: arXiv:1703.09312

[25] Jeffrey Mahler, Matthew Matl, Xinyu Liu, Albert Li, David Gealy, and Ken Goldberg, "Dex-Net 3.0: Computing robust vacuum suction grasp targets in point clouds using a new analytic model and deep learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.

[26] Ryo Matsumura, Kensuke Harada, Yukiyasu Domae, and Weiwei Wan, "Learning based industrial bin-picking trained with approximate physics simulator," in *International Conference on Intelligent Autonomous Systems*. Springer International Publishing, 2018, pp. 786–798.

[27] Rico Jonschkowski, Clemens Eppner, Sebastian Höfer, Roberto Martín-Martín, and Oliver Brock, "Probabilistic multi-class segmentation for the Amazon Picking Challenge," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 1–7.

[28] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[29] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You Only Look Once: Unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[30] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg, "SSD: Single shot multibox detector," in *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016, pp. 21–37.

[31] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, February 1992, pp. 239–256.

[32] Ming-Yu Liu, Oncel Tuzel, Ashok Veeraraghavan, and Rama Chellappa, "Fast directional chamfer matching," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 1696–1703.

[33] Bertram Drost, Markus Ulrich, Nassir Navab, and Slobodan Ilic, "Model globally, match locally: Efficient and robust 3D object recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 998–1005.

[34] Changhyun Choi, Yuichi Taguchi, Oncel Tuzel, Ming-Yu Liu, and Srikumar Ramalingam, "Voting-based pose estimation for robotic assembly using a 3D sensor," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 1724–1731.

[35] Takuya Torii and Manabu Hashimoto, "Model-less estimation method for robot grasping parameters using 3D shape primitive approximation," in *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2018, pp. 580–585.

[36] Andy Zeng, Kuan-Ting Yu, Shuran Song, Daniel Suo, Ed Walker, Alberto Rodriguez, and Jianxiong Xiao, "Multi-view self-supervised deep learning for 6D pose estimation in the Amazon Picking Challenge," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 1386–1383.

[37] Andy Zeng, Shuran Song, Johnny Lee, Alberto Rodriguez, and Thomas Funkhouser, "Tossingbot: Learning to throw arbitrary objects with residual physics," *arXiv preprint arXiv:1903.11239*, 2019.

[38] J. Bian, W. Lin, Y. Matsushita, S. Yeung, T. Nguyen, and M. Cheng, "Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 2828–2837.

[39] Gustavo Alfonso Garcia Ricardez, Felix von Drigalski, Lotfi El Hafi, Ming Ding, Jun Takamatsu, and Tsukasa Ogasawara, "Lessons from the Airbus Shopfloor Challenge 2016 and the Amazon Robotics Challenge 2017," in *18th SICE System Integration Division Annual Conf.*, Sendai, Japan, dec 2017, pp. 572–575.

[40] Gustavo Alfonso Garcia Ricardez, Felix von Drigalski, Lotfi El Hafi, Seigo Okada, Pin-Chu Yang, Wataru Yamazaki, Viktor Hoerig, Arnaud Delmotte, Akishige Yuguchi, Marcus Gall, Chika Shiogama, Kenta Toyoshima, Pedro Miguel Uriguen Eljuri, Rodrigo Elizalde Zapata, Ming Ding, Jun Takamatsu, and Tsukasa Ogasawara, "Warehouse picking automation system with learning- and feature-based object recognition and grasping point estimation," in *18th SICE System Integration Division Annual Conf.*, Sendai, Japan, dec 2017, pp. 2249–2253.