# Adaptive Perception and Manipulation for Autonomous Robotic Kitting in Dynamic Warehouse Environments

Møller, Philip Lund; Jami, Milad; Hansen, Karl Damkjær; Chrysostomou, Dimitrios

*Published in:*
IST 2023 - IEEE International Conference on Imaging Systems and Techniques, Proceedings

*DOI (link to publication from Publisher):*
10.1109/IST59124.2023.10355656



*Publication date:*
2023

*Document Version*
Accepted author manuscript, peer reviewed version

# Adaptive Perception and Manipulation for Autonomous Robotic Kitting in Dynamic Warehouse Environments

Philip Lund Møller
*DMD, Robotics & OT*
*Novo Nordisk A/S*
Hillerød, Denmark
pplm@novonordisk.com

Milad Jami
*DMD, Robotics & OT*
*Novo Nordisk A/S*
Hillerød, Denmark
mjam@novonordisk.com

Karl Damkjær Hansen
*Dept. of Materials & Production*
*Aalborg University*
Aalborg, Denmark
kdh@mp.aau.dk

Dimitrios Chrysostomou
*Dept. of Materials & Production*
*Aalborg University*
Aalborg, Denmark
dimi@mp.aau.dk

***Abstract*—Manual batch kitting in pharmaceutical manufacturing is labor intensive and prone to quality issues. Automation can improve productivity and reduce risks, but handling dynamic warehouse environments is challenging. This paper studies the feasibility of integrating autonomous industrial mobile manipulators (AIMMs) to enable flexible kitting without infrastructure changes. We propose an adaptive perception and manipulation system for localizing and grasping pharmaceutical containers from variable rack placements. Fiducial markers provide approximate navigation for the mobile base, then point cloud processing precisely estimates container poses for manipulation even under tight spacing and occlusion constraints. The key technical contribution is a segmentation approach leveraging the viewpoint to isolate the target container's front plane for robust localization. The proposed system was evaluated in a mock pharmaceutical kitting environment where it localized containers with an average translation and orientation error of respectively 6mm and 1.6 degree. Finally, it achieved a success rate of 96.66% for the kitting of individual containers.**

***Index Terms*—robotic kitting, autonomous mobile robots, point cloud processing, pose estimation**

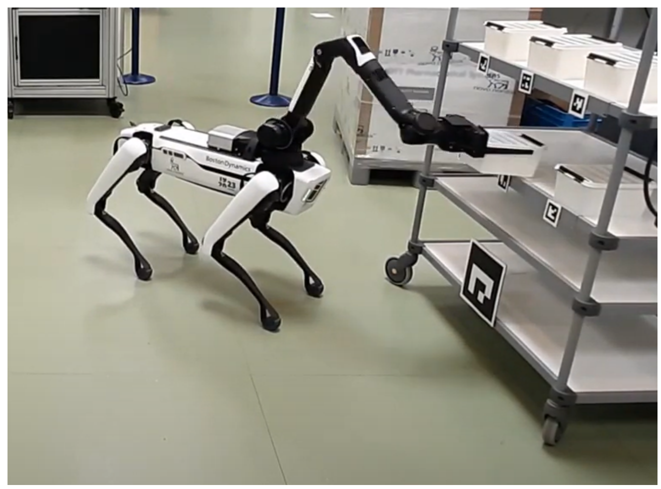

Fig. 1. Overview of the experimental setup where the Spot robot collects a container with injection device components.

## I. Introduction

The process of kitting has traditionally been highly manual due to the dynamic conditions it is subject to, both in terms of the process parameters, e.g., number of components and combinations, as well as the environment where it takes place [1]. Kitting areas generally feature various dynamic aspects such as human and forklift traffic, differing storage rack types, and containers with varying appearances, shapes and sizes, all of which must be handled [2].

However, most prior kitting research has focused on structured bins rather than dynamic environments. Many challenges persist in flexibly perceiving and manipulating objects from variable rack storage [3]. This problem is especially pertinent in pharmaceutical manufacturing, where kitting provides critical quality control but incurs high costs from stringent regulations if done manually [4].

The work presented in this paper was carried out at the premises of the pharmaceutical company Novo Nordisk, where kitting of production batches is one of many processes used to ensure product quality. Components can thereby be stored in sealed containers right up until the point where they are fed into production areas, while the constraint of one kit per batch simplifies line clearance and helps reduce the risk of cross-contamination. Therefore, heavy emphasis is placed on documentation and traceability of individual process steps, often done through scanning of barcodes and manual inputs to various systems, which prolongs the overall process and introduces many sources of error. As a result, automation of such kitting processes is a necessity.

Automation of kitting processes within traditional storage areas, without requiring extensive and costly restructuring, requires a complex system that is capable of intelligently handling various problems related to navigation, perception and manipulation [5]. Autonomous industrial mobile manipulators (AIMMs) have become indispensable for automating kitting operations in industrial environments, as they integrate perception, navigation and manipulation capabilities on a system usually comprised with a mobile platform and an industrial robotic manipulator [6]–[8].

In this work, we study the feasibility of integrating a Spot robot from Boston Dynamics (Fig. 1) for autonomous kitting processes specifically for pharmaceutical production settings by means of a tailored perception and pose estimation system. In particular, we examine the complex requirements surrounding accurate perception and dexterous manipulation in cluttered, unstructured environments. To robustly handle occlusion and variability during localization of arbitrarily placed containers for grasping, we propose a system which combines fiducial markers for approximate navigation with point cloud processing for precision pose estimation. The main contributions of this work are:

- Adaptive localization of pharmaceutical containers under tight spacing and occlusion constraints.
- Integration of this approach with fiducial marker strategies to avoid expensive infrastructure modifications.
- Experimental validation of autonomous kitting for streamlining a traditionally manual pharmaceutical process.

## II. Background

### A. Kitting use case at Novo Nordisk

The process of kitting for production batches at Novo Nordisk consists of the collection of a number of sealed containers that must be pulled out from storage racks in a designated kitting area, carried to, and placed on a pallet or trolley. Containers are located by a storage rack ID and an accompanying shelf location, defined as a row-column on the rack, both of which are indicated by markers and barcodes. The containers are placed freely within the approximate area above a shelf location marker, and thus, their exact position and orientation cannot be determined prior to immediate handling. It is therefore sufficient for Spot to approximately reach the area where the container is expected to be, however, an accurate 6D pose estimate for the desired container must be computed upon arrival.

### B. Localization with occlusion constraints

A core challenge in automating Novo Nordisk's kitting process is robustly localizing containers in dynamic environments where occlusions can create challenges for determine the right position of boxes in the shelves. While model-based techniques like RANSAC [9] have been applied in kitting operations before, assumptions about shelf geometry and known geometric primitives in the environment are still needed. Newer learning approaches are promising handling of incomplete point clouds in dynamic scenes [10] where advanced neural nets are used to complete shapes from partial inputs for grasp pose detection. However, large datasets are often required to train such neural networks and they pose a great challenge in safety-critical domains such as a pharmaceutical warehouse where validation of leaned models is difficult [11].

### C. Adaptive perception based on point cloud segmentation

Point cloud processing offers a robust solution for handling unstructured storage environments, with basic geometric methods such as planar segmentation [12] providing resilience against missing data due to occlusions. In a recent example, Liu et al. [13] have used region growing approaches to extract planar surfaces from noisy point clouds for industrial applications, and combined RANSAC with Euclidean clustering to enable robot pick and place in cluttered environments by extracting key surfaces.

Recent studies have also explored learning-based perception directly from raw point clouds, such as PointNet [14] and GndNet [15], containing neural network architectures that utilize unordered point sets for object classification and segmentation. However, the same challenge with the localization feature arises as large labeled datasets are required for training, which is a major challenge in industrial settings.

### D. Autonomous mobile manipulators in kitting operations

As discussed before, automating kitting operations in warehouses has been an active area of robotics and vision research [5], [16]. Early efforts focused on structured environments with known storage locations like bins and racks. Martinez et al. [17] developed a system for bin picking using a fixed robotic arm and a standard gripper which could recognize industrial parts in random positions. In similar fashion, Olesen et al. [18] enabled a robotic system to utilize deep learning policies for the detection of the parts and a multi-gripper switching strategy to efficiently grasp them. AIMMs such as Little Helper [19] and its dual arm alternatives [20] integrated autonomous navigation, flexible perception and adaptive manipulation in complex industrial environments for kitting and logistics operations.

The key advantage of AIMMs is that they do not restrict automation to predefined workcells but can flexibly operate throughout warehouses. This kind of mobility enables accessing items from varied locations and vantage points. This work in this paper builds upon the strengths of AIMMs for automating pharmaceutical kitting in dynamic storage environments by combining an occlusion-resilient localization method with the flexibility of point cloud segmentation

## III. Methodology

For the localization of individual containers, it is assumed that the end-effector of the manipulator can be aligned with and positioned directly in front of respective shelf location markers to achieve a consistent field of view for containers.

### A. Segmentation of point cloud

To estimate the pose of a container, its front plane must be isolated from the point cloud. The scene is segmented by iteratively fitting planes with RANSAC [21], removing inliers, and extracting significant planes above a threshold. This identifies key surfaces while filtering noise and the planes are further filtered to remove intersections. As the end-effector must center on the front plane, it is identified by orientation

Fig. 2. Iterative execution of RANSAC and removal of inliers can segment a large point cloud, such as that seen in the middle, into smaller planes, such as those seen to the right, while also removing overall noise.

and proximity to the shelf front. The extracted significant planes are then filtered by a k-nearest neighbor algorithm to remove potential points from intersections with other planes.

The camera's central position enables identifying the container's front plane. Firstly, the orientation of a plane relative to the camera frame is used to filter out the planes belonging to e.g., the shelf where the container sits on or the bottom of the shelf above it, by removing horizontal planes. Of the vertical planes, it can be assumed that the one closest to the camera will be that of the shelf front carrying the shelf location marker. The remaining are scored by size and proximity to select the largest nearest plane as the container front. This leverages the camera viewpoint to reliably segment the target container from cluttered storage. The segmentation process can be seen visualized in Fig. 2

### B. Estimation of grasp pose

From the plane defining the front of the container, a grasp pose can be computed that is centered around the top lip of the containers. The orientation of the end-effector at the grasp pose will be normal to the front plane, and thus, only a 2D offset, along the X- and Y-axis of this plane, must be computed for the grasp pose.

A bounding box and convex hull extract the quadrilateral shape of the container front. The four points from the convex hull that are closest to the corners of the bounding box are then assumed to be the true corners of the container front. The 2D offset for the grasp pose can then be computed as the halfway point between the two top corners of the quadrilateral. This whole process can be seen visualized in Fig. 3.

## IV. System Overview

Being able to localize an arbitrarily positioned container from immediately in front of its respective shelf location marker, the remaining system must support Spot in reaching this pose to pick the container and return it to the trolley. This requires navigation to and from the trolley and storage racks, along with an alignment procedure at the storage rack to place the end-effector in position for localizing the container.

### A. Hardware setup

Spot features a total of six cameras, five of which are placed around the body while the last one is placed inside the end-effector. Each body vision system features a projected IR stereo camera for depth images and a separate camera for greyscale images. The end-effector vision system also features a projected IR stereo camera but instead integrates a separate 4K RGB camera for color images.

A custom end-effector was developed for Spot as the default claw gripper was not suitable for handling containers. This end-effector is configured to mechanically interface with the motors of the existing end-effector thereby replacing the claw with a parallel mechanism featuring two prongs to provide vertical support around the lip of containers during grasping and allow for slipping in between tightly packed containers. The prongs are chamfered towards the center of the end-effector to push containers against an overhang to clutch them.

### B. Navigation

The requirement to document kitting processes at Novo Nordisk through scanning of barcodes, presents an opportunity for combining problems related to navigation with documentation by adopting a fiducial marker identification system. This also serves as a better interface for a mobile vision system such as Spot. Individual sets of markers are used to identify respectively storage racks and individual shelf locations within the given storage rack. The pose of fiducial markers, such as ArUco markers [22], is localized relative to a camera from single-shot images, assuming intrinsic parameters are known, while simultaneously conveying binary information through its internal content. This content can thus be logged by Spot and images can be stored to document the steps of the kitting process, while the poses of the marker itself can be used to guide Spot towards a desired storage rack through relatively simple logic and point-to-point path planning.

Generally, pose estimates for fiducial markers becomes more accurate at closer ranges, and therefore the pose estimate will be refined as Spot approaches a desired target. The approach will bring Spot towards the target at an angle to mitigate rotational ambiguities during pose estimation [23], and once an estimate from a sufficiently close distance is made, Spot will turn to face the target head-on at a 1m offset directly in front of the marker.

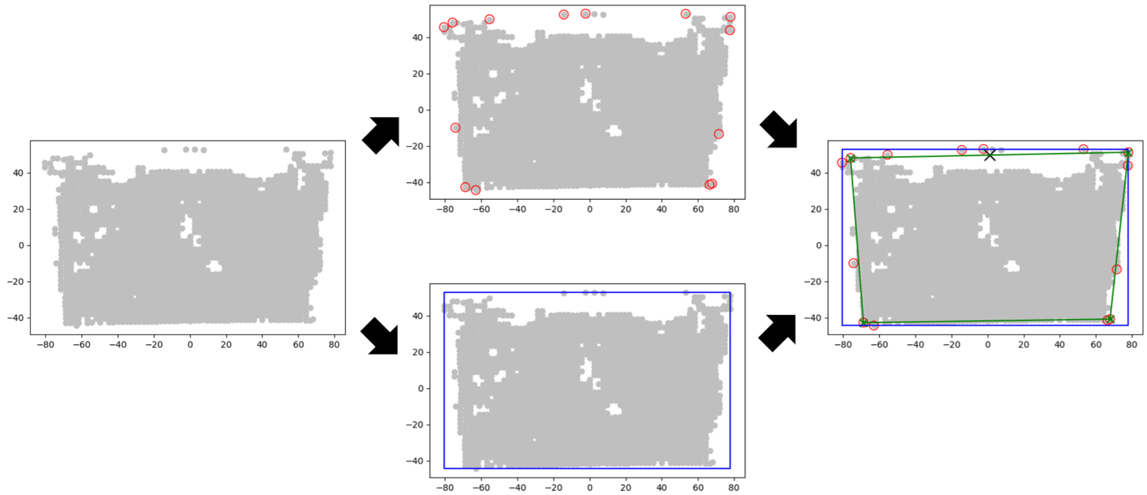

Fig. 3. Estimation of container front shape. To the left, the container front can be seen projected to 2D. In the middle, a convex hull (red) and a bounding box (blue) is fitted around this data, and finally combined to the right to determine the overall shape (green). The grasping pose can seen marked by a black cross to the right.

### C. Alignment to shelf location

From this position, an overview image is taken to establish the present shelf locations. The height of the desired shelf location marker, relative to the ground, is then determined, and Spot is moved to position the base of its manipulator directly in front of the marker, along its Z-axis, with the end-effector raised to the height of it to ensure it will be visible from the end-effector camera. The pose is then estimated for the marker and the end-effector is aligned with it at a 0.3m offset along its Z-axis. From this pose, Spot is aligned and ready to localize a given container at the shelf location.

Simple linear trajectories are then computed for the end-effector to pick the container after localization, which, beyond the estimated grasp pose, will cycle through an approach pose, a lift pose, and a retract pose, all of which are computed as offsets from the grasp pose. Force feedback is used while the end-effector closes to verify and log whether a grasp is successful.

### D. Alignment to trolley

The trolley has a fiducial marker to aid in navigation and documentation, and therefore, an almost identical approach is used for aligning Spot to the trolley before placing the container on it. However, here the desired pose for the container cannot be indicated by another fiducial marker, but must instead be governed by another system. For this integration, a simple pre-defined grid layout was used to show the potential for accurately stacking containers during the kitting.

### E. Error handling

To further help mitigate the influence of dynamic conditions during various localizations, a global philosophy is adopted for error handling which permits Spot to fail an objective once and try again, however, failing twice means that an error should be reported. Some recursive behaviour is therefore triggered in these cases to retry some of the tasks up until the previous point of failure. The logic behind this error handling can be seen mapped out in Figure 4.

### F. Safety layer

Dynamic obstacles within the warehouse environment must still be considered, as particularly human traffic is likely to occur and their safety must be assured. Some level of obstacle avoidance is included within Spot off-the-shelf, however, the potential unpredictable behaviour associated with this may pose a safety hazard in itself. Therefore, an additional safety layer is integrated which will cause Spot to stop moving whenever a human comes closer than 2m. This safety layer leverages YOLOv5 [24] to recognize humans in images sampled from each body camera continuously, and, if a human is detected, will then measure the distance to various points from a region of interest within the fitted bounding box around the human.

## V. Experiments

To demonstrate the performance of the proposed kitting system, we performed two experiments: one to determine how accurately it can localize desired containers, and another to determine how successfully it can complete a kitting. These experiments were carried out in a mock kitting area setup (Fig. 5), which consists of: a wheeled storage rack featuring six shelf locations, a wheeled trolley used in the current kitting process, one set of 15cm ArUco markers for storage racks and trolley and one set of 5cm ArUco markers for shelf locations.

### A. Container Localization Accuracy

This experiment estimates the accuracy by calibrating a fixed trajectory for Spot to place a container on the storage rack and store its pose in the world frame as a ground truth. Adhesive material is added to the bottom of the container to prevent movement during release. Spot then approaches

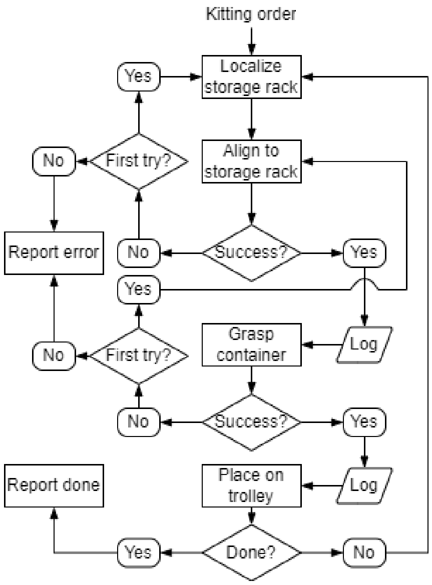


Fig. 4. Error handling is included whenever the robot fails to align at a desired position or when it fails to grasp a container.

Fig. 5. Overview of experimental setup. The trolley and storage rack are indicated with large ArUco markers, respectively left and right, while shelf locations for each container within the storage rack are indicated with small ArUco markers.

the storage rack five times to localize the container. This test is carried out for three different calibrated trajectories. The results of each test can be seen summarized in Table I. Translation error is defined as the Euclidean distance from the computed translation to the ground truth translation while orientation error is computed as the angle of the angle-axis representation of the quaternion difference between the ground truth quaternion and a given test quaternion. The difference quaternion, $q_d$, is found through:

$$q_d = q_2 q_1^{-1} \tag{1}$$

Where $q_2$ is the ground truth quaternion and $q_1$ is a given test quaternion. From the resulting quaternion, the angle is recovered through:

$$\theta = 2\ \text{atan2}\left(\sqrt{q_{dx}^2 + q_{dy}^2 + q_{dz}^2},\ q_{dw}\right) \tag{2}$$

where $q_{dx}$, $q_{dy}$, $q_{dz}$ and $q_{dw}$ are the individual elements of the difference quaternion.

From Table I, it can be deduced that the localization algorithm achieved an average translation error of 0.006m, or 6mm, and an average orientation error of 1.6 degree.

TABLE I
CONTAINER LOCALIZATION ACCURACY

| | **No.** | **Translation error (m)** | **Orientation error (deg)** |
|---|---|---|---|
| **Experiment 1** | *1* | 0.0058 | 0.3329 |
| | *2* | 0.0112 | 0.3320 |
| | *3* | 0.0151 | 0.3026 |
| | *4* | 0.0074 | 0.3060 |
| | *5* | 0.0058 | 0.3139 |
| **Experiment 2** | *6* | 0.0075 | 2.2610 |
| | *7* | 0.0024 | 2.2845 |
| | *8* | 0.0067 | 2.2643 |
| | *9* | 0.0037 | 2.2724 |
| | *10* | 0.0036 | 2.2887 |
| **Experiment 3** | *11* | 0.0041 | 2.2617 |
| | *12* | 0.0035 | 2.2696 |
| | *13* | 0.0041 | 2.2785 |
| | *14* | 0.0058 | 2.2932 |
| | *15* | 0.0037 | 2.2927 |

### *B. Kitting Success Rate*

This experiment will evaluate the overall robustness of the kitting system by having Spot attempt to complete ten full kitting processes, each of which consists of six containers. Kitting of a given container is considered successful if it has been picked from the storage rack and moved to the trolley, however, failing to kit a single container means that the overall kitting operation was not successful.

Results from these tests are shown in Table II. It can be deduced that the kitting success rate of individual containers is 96.66% on average, however, as two overall kitting operations failed their success rate is only 80%.

TABLE II
SUCCESS RATE OF THE KITTING OPERATION

| **No.** | **Success/fail** | | | | | | **Rate** |
|---|---|---|---|---|---|---|---|
| **1** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| **2** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| **3** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| **4** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| **5** | ✓ | ✓ | ✓ | ✓ | × | ✓ | 83.33% |
| **6** | ✓ | ✓ | ✓ | ✓ | ✓ | × | 83.33% |
| **7** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| **8** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| **9** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| **10** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |

## VI. DISCUSSION & FUTURE WORK

The experimental results validate the proposed system can reliably automate pharmaceutical kitting operations. The integration of mobility, perception, and manipulation technologies enabled flexible pick and place and automatic kitting in dynamic warehouse environments in a pharmaceutical context. Additionally, plane segmentation of point cloud data demonstrated highly accurate localization of containers despite tight spacing and occlusion on shelves.

Approximate positioning using fiducial markers proved effective for the navigation of Spot scoring an accuracy within 6mm and 1.6 degrees which is on par with state-of-the-art results for warehouse kitting operations. While system performance exceeded 95% container-wise success, overall kitting completion was 80% due to plane segmentation failures. This highlights limitations in the perceptual pipeline which could be improved with the application of a more selective feature extraction methodology and higher resolution imaging sensors.

The current system performs with high success rate in a small warehouse area. Naturally, larger-scale replication could reveal additional challenges. Factors like sensor noise and calibration errors emerge in long-term robotic deployments so robust mechanisms for error detection and recovery will be important for maintaining the desired reliability and robustness. Enhanced planning and control methods could also make the system more reactive when failures occur.

## VII. Conclusion

This paper presents an AIMM for pharmaceutical kitting automation. The robot leverages fiducial markers for approximate navigation and point cloud processing to accurately localize arbitrarily placed containers. This facilitates reliable grasping and kitting through integration of mobility, perception, and manipulation technologies. Extensive experiments with our system validated system performance where the robot achieved highly precise 6D pose estimation within 6mm and 1.6 degrees for successful container localization and grasping exceeding 96%. Tests in a mock warehouse environment emulated real-world challenges like cluttered shelves and dynamic environments.

The results comprehensively validate the feasibility of using AIMMs to provide automation benefits in pharmaceutical facilities without expensive infrastructure changes. In particular, plane segmentation enabled robust perception despite tight container spacing and complex occlusions. The integrated system adapts to variable container placements and confirms that the key technical challenges surrounding unstructured navigation, perception, and manipulation in automating pharmaceutical kitting can be addressed successfully. In conclusion, this work establishes AIMMs as a viable approach for increasing quality and productivity in pharmaceutical manufacturing by automating critical warehousing functions such as kitting with high success rate.

## References


[1] Y. A. Bozer and L. F. McGinnis, "Kitting versus line stocking: A conceptual framework and a descriptive model," *International Journal of Production Economics*, vol. 28, no. 1, pp. 1–19, 1992.

[2] M. E. A. Boudella, E. Sahin, and Y. Dallery, "Kitting optimisation in just-in-time mixed-model assembly lines: assigning parts to pickers in a hybrid robot–operator kitting system," *International Journal of Production Research*, vol. 56, no. 16, pp. 5475–5494, 2018.

[3] A. C. Caputo, P. M. Pelagagge, and P. Salini, "A model for planning and economic comparison of manual and automated kitting systems," *International Journal of Production Research*, vol. 59, no. 3, pp. 885–908, 2021.

[4] R. Hanson and A. Brolin, "A comparison of kitting and continuous supply in in-plant materials supply," *International Journal of Production Research*, vol. 51, no. 4, pp. 979–992, 2013.

[5] S. Balakirsky, Z. Kootbally, T. Kramer, A. Pietromartire, C. Schlenoff, and S. Gupta, "Knowledge driven robotics for kitting applications," *Robotics and Autonomous Systems*, vol. 61, no. 11, pp. 1205–1214, 2013.

[6] M. Hvilshøj, S. Bøgh, O. Skov Nielsen, and O. Madsen, "Autonomous industrial mobile manipulation (aimm): past, present and future," *Industrial Robot: An International Journal*, vol. 39, no. 2, pp. 120–135, 2012.

[7] D. Holz, A. Topalidou-Kyniazopoulou, F. Rovida, M. R. Pedersen, V. Krüger, and S. Behnke, "A skill-based system for object perception and manipulation for automating kitting tasks," in *2015 IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA)*. IEEE, 2015, pp. 1–9.

[8] C. Li, A. K. Hansen, D. Chrysostomou, S. Bøgh, and O. Madsen, "Bringing a natural language-enabled virtual assistant to industrial mobile robots for learning, training and assistance of manufacturing tasks," in *2022 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2022, pp. 238–243.

[9] D. Pavlichenko, G. M. García, S. Koo, and S. Behnke, "Kittingbot: A mobile manipulation robot for collaborative kitting in automotive logistics," in *Intelligent Autonomous Systems 15: Proceedings of the 15th International Conference IAS-15*. Springer, 2019, pp. 849–864.

[10] W. Qin, Q. Hu, Z. Zhuang, H. Huang, X. Zhu, and L. Han, "Ippe-pcr: a novel 6d pose estimation method based on point cloud repair for textureless and occluded industrial parts," *Journal of Intelligent Manufacturing*, vol. 34, no. 6, pp. 2797–2807, 2023.

[11] N. S. Arden, A. C. Fisher, K. Tyner, X. Y. Lawrence, S. L. Lee, and M. Kopcha, "Industry 4.0 for pharmaceutical manufacturing: Preparing for the smart factories of the future," *International Journal of Pharmaceutics*, vol. 602, p. 120554, 2021.

[12] L. Li, F. Yang, H. Zhu, D. Li, Y. Li, and L. Tang, "An improved ransac for 3d point cloud plane segmentation based on normal distribution transformation cells," *Remote Sensing*, vol. 9, no. 5, p. 433, 2017.

[13] H. Liu, R. Song, X. Zhang, and H. Liu, "Point cloud segmentation based on euclidean clustering and multi-plane extraction in rugged field," *Measurement Science and Technology*, vol. 32, no. 9, p. 095106, 2021.

[14] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

[15] A. Paigwar, Ö. Erkent, D. Sierra-Gonzalez, and C. Laugier, "Gndnet: Fast ground plane estimation and point cloud segmentation for autonomous vehicles," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2150–2156.

[16] M. Mancini, H. Karaoguz, E. Ricci, P. Jensfelt, and B. Caputo, "Kitting in the wild through online domain adaptation," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1103–1109.

[17] C. Martinez, R. Boca, B. Zhang, H. Chen, and S. Nidamarthi, "Automated bin picking system for randomly located industrial parts," in *2015 IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*. IEEE, 2015, pp. 1–6.

[18] A. S. Olesen, B. B. Gergaly, E. A. Ryberg, M. R. Thomsen, and D. Chrysostomou, "A collaborative robot cell for random bin-picking based on deep learning policies and a multi-gripper switching strategy," *Procedia Manufacturing*, vol. 51, pp. 3–10, 2020.

[19] M. Hvilshøj and S. Bøgh, ""little helper"—an autonomous industrial mobile manipulator concept," *International Journal of Advanced Robotic Systems*, vol. 8, no. 2, p. 15, 2011.

[20] J. F. Buhl, R. Grønhøj, J. K. Jørgensen, G. Mateus, D. Pinto, J. K. Sørensen, S. Bøgh, and D. Chrysostomou, "A dual-arm collaborative robot system for the smart factories of the future," *Procedia manufacturing*, vol. 38, pp. 333–340, 2019.

[21] G. Kahn, P. Sujan, S. Patil, S. Bopardikar, J. Ryde, K. Goldberg, and P. Abbeel, "Active exploration using trajectory optimization for robotic grasping in the presence of occlusions," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 4783–4790.

[22] M. Kalaitzakis, B. Cain, S. Carroll, A. Ambrosi, C. Whitehead, and N. Vitzilaios, "Fiducial markers for pose estimation: Overview, applications and experimental comparison of the artag, apriltag, aruco and stag markers," *Journal of Intelligent & Robotic Systems*, vol. 101, pp. 1–26, 2021.

[23] G. Schweighofer and A. Pinz, "Robust pose estimation from a planar target," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2024–2030, 2006.

[24] G. Jocher, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Real-time Instance Segmentation," Nov. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7347926