Toward Future Automatic Warehouses: An Autonomous Depalletizing System Based on Mobile Manipulation and 3D Perception

Jacopo Aleotti [1,*], Alberto Baldassarri [2], Marcello Bonfè [3], Marco Carricato [2], Davide Chiaravalli [4], Roberto Di Leva [2], Cesare Fantuzzi [5], Saverio Farsoni [3], Gino Innero [2], Dario Lodi Rizzini [1], Claudio Melchiorri [4], Riccardo Monica [6], Gianluca Palli [4], Jacopo Rizzi [3], Lorenzo Sabattini [5], Gianluca Sampietro [5] and Federico Zaccaria [2]

[1] DIA—Department of Engineering and Architecture, University of Parma, 43124 Parma, Italy; dlr@ce.unipr.it
[2] DIN—Department of Industrial Engineering, University of Bologna, 40136 Bologna, Italy; alberto.baldassarri@unibo.it (A.B.); marco.carricato@unibo.it (M.C.); roberto.dileva@unibo.it (R.D.L.); gino.innero2@unibo.it (G.I.); federico.zaccaria3@unibo.it (F.Z.)
[3] Department of Engineering, University of Ferrara, 44122 Ferrara, Italy; marcello.bonfe@unife.it (M.B.); saverio.farsoni@unife.it (S.F.); jacopo.rizzi@unife.it (J.R.)
[4] DEI—Department of Electrical, Electronic and Information Engineering, University of Bologna, 40136 Bologna, Italy; davide.chiaravalli2@unibo.it (D.C.); claudio.melchiorri@unibo.it (C.M.); gianluca.palli@unibo.it (G.P.)
[5] DISMI—Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, 42122 Reggio Emilia, Italy; cesare.fantuzzi@unimore.it (C.F.); lorenzo.sabattini@unimore.it (L.S.); gianluca.sampietro@unimore.it (G.S.)
[6] CIDEA—Center for Energy and Environment, University of Parma, 43124 Parma, Italy; riccardo.monica@unipr.it
[*] Correspondence: jacopo.aleotti@unipr.it

Abstract: This paper presents a mobile manipulation platform designed for autonomous depalletizing tasks. The proposed solution integrates machine vision, control and mechanical components to increase flexibility and ease of deployment in industrial environments such as warehouses. A collaborative robot mounted on a mobile base is proposed, equipped with a simple manipulation tool and a 3D in-hand vision system that detects parcel boxes on a pallet, and that pulls them one by one on the mobile base for transportation. The robot setup allows to avoid the cumbersome implementation of pick-and-place operations, since it does not require lifting the boxes. The 3D vision system is used to provide an initial estimation of the pose of the boxes on the top layer of the pallet, and to accurately detect the separation between the boxes for manipulation. Force measurement provided by the robot together with admittance control are exploited to verify the correct execution of the manipulation task. The proposed system was implemented and tested in a simplified laboratory scenario and the results of experimental trials are reported.

Keywords: mobile manipulation; robot vision; industrial depalletizing

1. Introduction

Industrial automated warehouses are designed to optimize the transportation and distribution of goods usually stored in cardboard boxes [1]. In this work, we consider the task of automated depalletizing (unloading) a pallet containing homogeneous boxes, namely the origin pallet, with the purpose of composing a new pallet containing potentially different boxes, namely the mixed pallet, by collecting items from different origin pallets. Homogeneous pallets are organized as a grid of cardboard boxes of the same type arranged on multiple stacked layers.

Robotic depalletizers in automated warehouses are usually bulky and limited to four degrees of freedom (DOFs), requiring very high payloads to move large end-effectors.

Robot end-effectors are usually designed to pick up an entire layer of boxes from the origin pallet and place it on a distribution and serialization system composed of multiple conveyor belts, even in the case of light boxes or when only a few boxes are needed to be transferred from the origin pallet. Moreover, since more than ten thousand different types of goods may need to be stored in a single food and beverage factory, warehouses require large buildings for intermediate buffers to temporarily store the depalletized items before moving them to the destination (mixed) pallets.

In this paper, we address the automated depalletizing problem by defining a new system able to integrate safety, maneuverability and ease of interaction for a palletizing/depalletizing task performed in an industrial environment. This system could be considered a first prototype of a multi-purpose platform with high reconfiguration capabilities for the interaction and assistance of human operators in a shared warehouse environment. To this end, our system exploits a serial collaborative robot (cobot) mounted on an autonomous mobile robot (AMR). Both the serial arm and the AMR must be endowed with collaborative features, since the overall robotic system must be able to safely operate in an industrial environment shared with humans. The cobot is equipped with an eye-in-hand time of flight (ToF) 3D camera. The cobot, an UR10e from Universal Robots, has six DOFs, providing a wide workspace for box extraction, 10 Kg payload, and wrist force/torque sensor for tool wrench measurement. A shovel-shaped tool mounted on the cobot end-effector is used to fit in the gap between boxes and to pull the cardboard boxes one by one on the mobile base for transportation. The choice of this type of tool with respect to other standard solutions allows to reduce both the total execution time and the load on the manipulator. The 3D camera is located on the cobot end-effector to detect the pose of the boxes and the separation gap between them for tool insertion.

The design of the proposed mobile manipulation system was driven by the fact that payload is one of the main bottleneck problems faced by collaborative mobile robots in industrial environments. Indeed, due to safety and maneuverability, only small cobots can be placed on AMRs. Therefore, the proposed solution is based on decoupling object displacement and lifting operations, leaving the former to the cobot and the latter to an automatic lifting device located on the AMR, next to the cobot. The cobot manipulates the item at hand (e.g., a box) without lifting it, but rather by dragging it onto the lifting device.

2. Related Work

The optimization of operations inside a warehouse is currently characterized by several research areas. In particular, the implementation of a robotic system able to assist human workers in their tasks requires the capability to effectively plan and optimize the sequence of actions [2–4], to correctly localize persons and objects [5] and to interact with the environment and other systems in the network [6]. Since the advent of Industry 4.0, the requirements of optimizing the transport and distribution of products has encouraged the use of autonomous guided vehicles within modern automated warehouses [7]. Moreover, autonomous mobile robots have been introduced to improve flexibility and robustness in modern industrial contexts [8]. Nowadays, the potential collaboration between cobots and mobile platforms has received increasing efforts from industrial and scientific research, particularly for logistic applications [9]. Bonini et al. [10] explored the possibility of employing a mobile collaborative system for palletizing operations. The hardware consists of a lifting device equipped with a pneumatic gripper, a powered conveyor and a set of sensors which allows interaction between the system itself and the environment. Vacuum grippers are usually adopted for depalletizing cardboard boxes [11–13]. However, this solution is not generally safe, since cardboard boxes cannot always hold the weight of the items that they contain. Nakamoto et al. [13] designed a depalletizing system where a second robot arm is used to support the boxes during motion. A robot manipulation system based on 3D vision and a vacuum gripper, for the detection and unloading of cardboard boxes from shipping containers or semi-truck trailers, is presented in [11]. A fixed-base robot depalletizing system designed for supermarket logistic processes is described in [12]. Similarly,

the application of vacuum technology was proposed in [13,14], whereas Matsuo et al. [15] developed a mobile robot equipped with a self-weight compensation system. However, these solutions may lack flexibility, and vacuum gripping solutions may be hard to implement on AMRs, especially small-sized ones. In contrast, with previous systems, a non-prehensile approach of manipulation was suggested by Lim et al. [16], where items are dragged aboard the mobile robotized system. Non-prehensile manipulation strategies are a subject of growing interest, both in the research and industrial fields, because with the reduction in the overall weight supported by the gripper, the minimization of the risk of falls and the capability to perform specific motions in a cluttered environment would be impossible under normal grasping scenarios. In [17], a new planning framework to exploit the funneling effect of pushing to deal with uncertain and clutter environments is proposed. Similarly, in [18] a framework that takes into account the human presence in cluttered environments is analyzed. Acharya et al. [19] presented a motion planning analysis for the optimization of the stability and control of an asymmetric object during non-prehensile manipulation. Ardakani et Al. [20] proposed a quasi-static analysis to define a dynamical system to predict the object behavior in function of friction forces. In the industrial scenario [21], the manipulation of a flexible belt exploiting the friction between the belt and the gripper is discussed. Our work aimed to promoting new studies on the interaction forces between the object to be picked and the support layer to further simplify the depalletizing operation in a shared human–robot environment.

In terms of perception, similarly to what is proposed in this paper, Hashimoto et al. [22] presented a genetic algorithm to recognize loads on a pallet. The main differences are that the method in [22] adopted a fixed gray scale camera placed above the pallet, while we adopted an eye-in-hand camera system and admittance control to simultaneously achieve the accurate positioning of the robot arm and a controlled interaction with the boxes. Yunardi et al. [23] investigated a method to determine the size of parcels moving on a conveyor belt using RGB cameras. Prasse et al. [24] proposed a system to detect the pose of parcels located on a pallet by combining a time of flight (ToF) sensor and RFIDs, which were used to compute the 3D structure of the layer. Katsoulas et al. [25] proposed methods for the recognition of arbitrary size boxes in cluttered environments using a planar laser scanner, mounted on a robot arm in eye-in-hand configuration. A drawback is that, since they are based on 2.5D edge detectors, they could fail to detect aligned boxes in contact with one another. In [26], a robotic depalletizing system was proposed using uncalibrated vision and 3D laser-assisted image analysis. All sensors were attached to the ceiling. In [27], an RGB-D vision system based on pattern matching was developed for the localization of heterogeneous cases in a depalletizing robotic cell.

3. Concept and Mechanical Design

The items to be handled are cardboard boxes placed on a standard Euro Pallet (EPAL), in such a manner that the parcel boxes form at most four vertical layers separated by rigid interlayers. The overall depalletizing task may be subdivided into the following sequential steps:

- Robot's self-localization within the workspace;
- Autonomous navigation towards the desired pallet;
- Pose detection of boxes on the top layer of the pallet;
- Extraction of boxes from the pallet and placement aboard the robot.

A manipulation strategy based on dragging goods aboard the mobile robot is employed, in order to overcome the limitations of standard grabbing manipulation, such as a limited robot payload or the possibility of handling only packages that are able to sustain their own weight.

We propose using a serial collaborative manipulator UR10e, provided by Universal Robots (www.universal-robots.com/products/ur10-robot/, accessed on 24 June 2021), installed on a Mir100 AMR, supplied by Mobile Industrial Robots (www.mobile-industrial-robots.com/en/solutions/robots/mir100/, accessed on 24 June 2021) as illustrated in

Figure 1. In addition, a scissor lifting mechanism is integrated to collect boxes on board. Thus, the displacement and lifting functions are decoupled. The top of the lifting mechanism is equipped with an idler-roller conveyor to allow items to be dragged and collected. During the manipulation phase, a swivel hatch, mounted on the terminal part of the conveyor, can be set in either open or closed configuration. The former position enables items to be dragged either from the pallet to the conveyor or vice versa, whereas the latter position prevents the boxes from falling out during the AMR navigation.
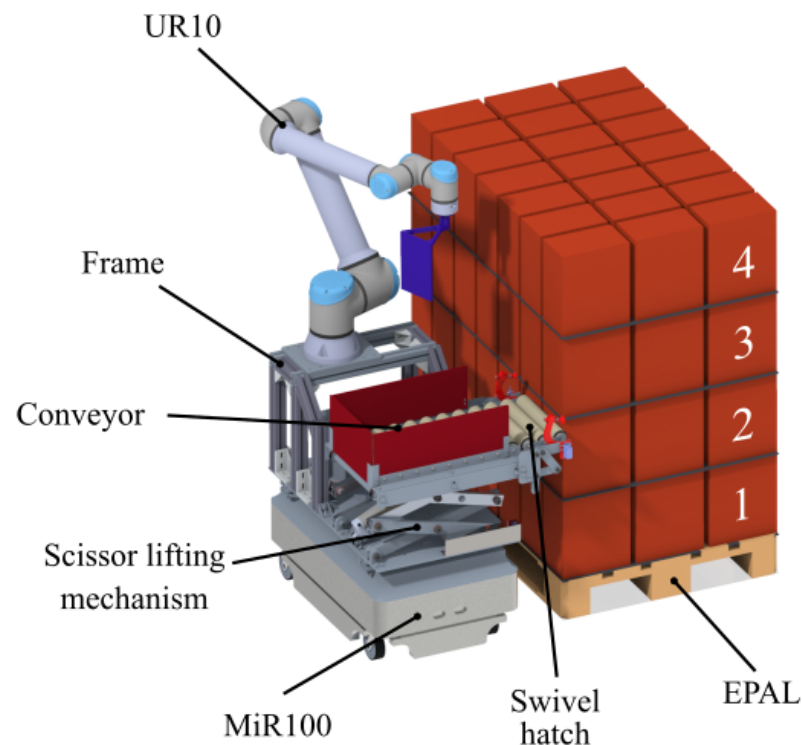

Figure 1. Overview of the mobile robot and its components.

The operational capability of the overall system is influenced by how the cobot and the lifting mechanism are integrated on the AMR. For this reason, three different solutions were analyzed. In the first solution, the cobot is installed on a fixed support, and the installation height is selected by optimizing the cobot workspace with respect to the box locations. In this way, system stability is not excessively penalized, but the fixed cobot location limits the boxes that can be reached. In the second solution, the cobot is installed on top of the lifting mechanism, thus simplifying task execution, since the cobot base is always aligned with the conveyor and the box layer to be handled. As a drawback, a more powerful lifting mechanism is needed. The last solution requires the installation of the cobot on a telescopic actuator, thus leading to independent movements of the cobot and the lifting mechanism. However, cost-effectiveness is penalized. Moreover, in both the second and the third solution, system stability decreases when handling the higher boxes. Another aspect to be considered when choosing the optimal hardware architecture is that, due to the rectangular footprint of the MiR AMR, the cobot and the lifting mechanism can be placed in either longitudinal or transverse arrangement. The former (Figure 2a) enhances stability, but possible interference between the lifting mechanism and the cobot limits the base joint rotation. On the other hand, the latter arrangement (Figure 2b) overcomes such limitations by virtue of the larger distance of the cobot with respect to the lifting mechanism, however, AMR navigation may be more challenging because of the non-omnidirectional MiR steering system. As a result of the design analysis (with the latter being thoroughly described in [28]), we chose the installation of the cobot on a fixed support with a transverse arrangement. It is worth observing that when the AMR is close to the pallet, at most two boxes can be processed, due to the limited workspace

of the robot arm; thus, handling the third box of the same row requires the AMR to reposition on the opposite side of the pallet. This drawback can be overcome by using a cobot with a larger workspace mounted on a larger AMR.

(**a**) Longitudinal layout.   (**b**) Transverse layout.
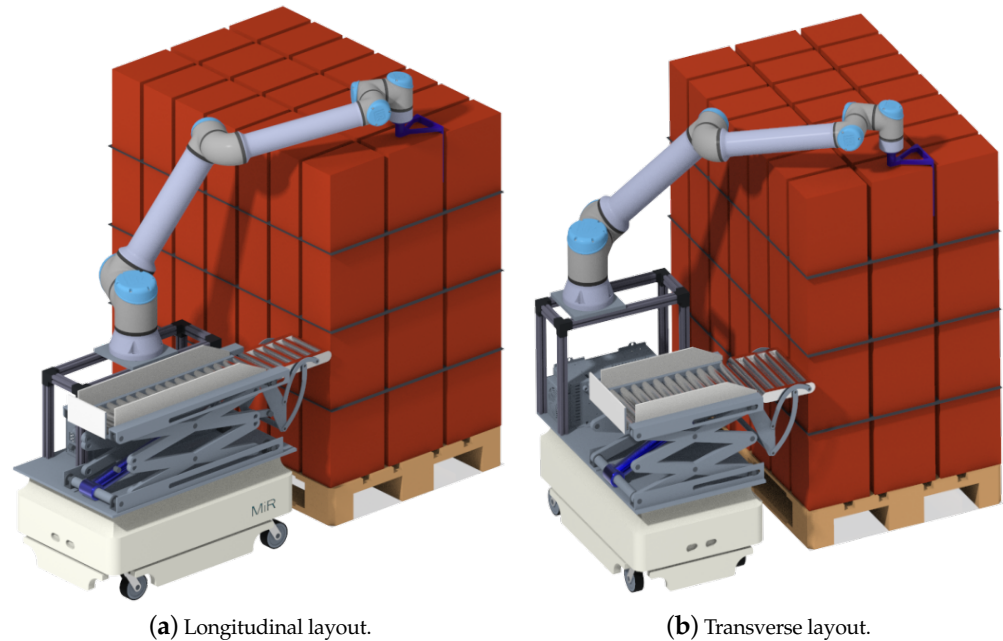
Figure 2. Comparison between transverse (**a**) and longitudinal (**b**) layout.

Concerning the lifting mechanism, it is composed of two scissor linkages placed in parallel and a linear electric actuator that acts on a transverse ledger. A four-bar linkage enables the rotational motion of the swivel hatch, as represented in Figure 3. For the four-bar actuation, we select a passive solution for the sake of simplicity, lightness and cost-effectiveness. By employing a crank-slider mechanism, the linear motion of the top wheel installed in the scissor lifting mechanism can be converted into an angular rotation of the four-bar crank.
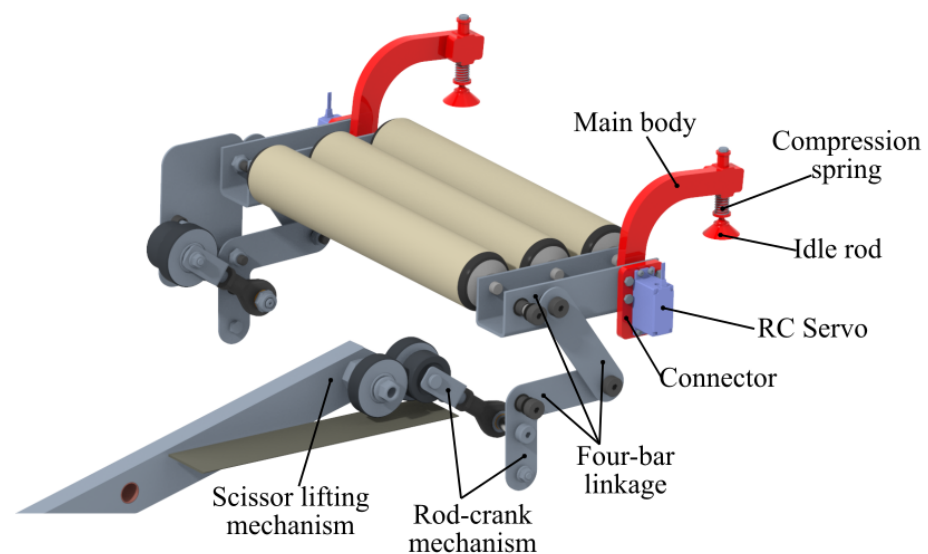
Figure 3. Swivel-hatch mechanism equipped with rotating clips.

Friction phenomena during box manipulation might result in an undesired sliding of the interlayer. Therefore, we use a pair of rotating clips (Figure 3) which keep the interlayer in place and prevent it from slipping. Each clip comprises a RC-Servo motor, a main body

rigidly attached to the motor, a sliding rod, and a compression spring. The RC-Servo actuates the rotation of the main body, resulting in a spring compression and a force applied on the interlayer.

4. Simulations

The industrial scenarios considered in this work were simulated using CoppeliaSim (www.coppeliarobotics.com/coppeliaSim, accessed on 24 June 2021) and evaluated in terms of task execution time and motion feasibility.

Simulations include a comparison between the dragging manipulation approach against a standard pick-and-place solution.

One layer of 21 palletized boxes (of size 250 × 150 × 300 mm, arranged in a grid of 7 rows, 3 columns) must be transported from an initial pallet to a storage pallet, maintaining the initial grid arrangement of the boxes. It is assumed that during the manipulation phase of the items the total payload attached to the robot (i.e., the tool and the box) is always lower than the UR10e maximum payload.

The first simulated scenario (Procedure 1) involves the usage of a shovel-shaped tool on the UR10e end-effector. The tool drags the item on the lifting device. In this case, it is assumed that the AMR can transport only two boxes per journey. Making such assumptions, the task can be accomplished in 11 journeys: for each of the seven rows of the box grid, the first two columns items are manipulated in the first seven journeys, then the items on the third column are handled in the remaining four journeys, with a quite different picking method. Indeed, a pair of items (in the same row) of the first two columns, once the AMR is positioned next to them, can be pulled by the manipulator without any additional motion of the AMR, whereas the AMR must be moved to manipulate a pair of items (in different rows) of the third column. Figure 4 shows the AMR/cobot system and the grid of 21 boxes highlighting the items involved in each of the 11 journeys.
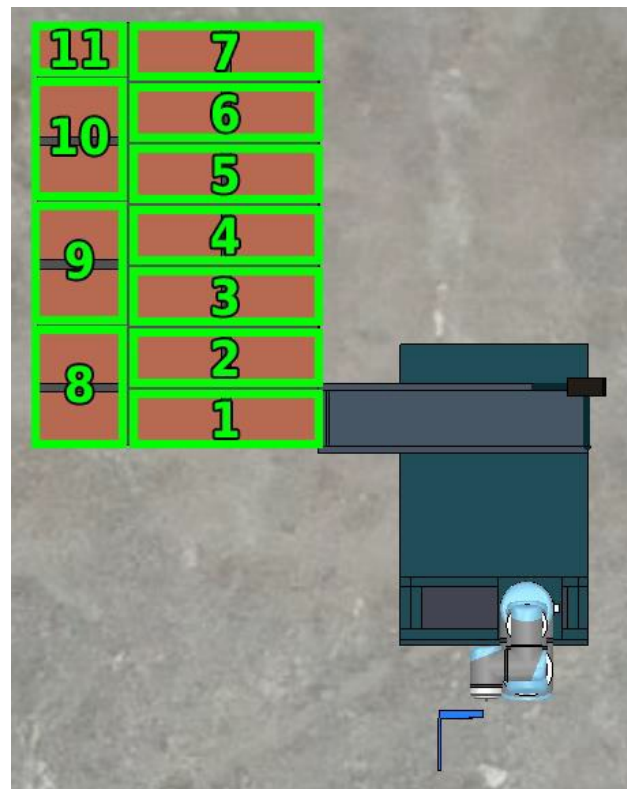

Figure 4. The grid of 21 boxes named according to the order of manipulation.

The shovel-shaped tool of the cobot allows to load a box on the lifting device by means of a linear dragging movement. Firstly, the manipulator executes a collision-free movement

from its current initial position towards the detected approaching frame between two consecutive boxes. As will be further described in Section 6, such a movement consists of different point-to-point motions, because of the need for taking a first 3D image of the pallet top layer and then a second image closer to the identified gap between two boxes to refine the detection accuracy. Afterwards, three linear motions are required to insert the tool into the detected gap, drag the box and finally bring the manipulator back to the next starting position.

Procedure 1 lists the consecutive motions that are required to perform the loading (and similarly, the unloading) of a single box on the AMR.

Procedure 1: one-box dragging motion sequence, shovel-shaped tool

---

1 Point-to-point motion to box target frame
2 Linear downward approaching movement (200 mm)
3 Linear dragging movement (600 mm)
4 Linear upward movement (300 mm)

The collision-free point-to-point motions are generated by means of the Open Motion Planning Library framework (OMPL) (https://ompl.kavrakilab.org/, accessed on 24 June 2021), exploiting the OMPL wrapper included into CoppeliaSim. OMPL is also integrated into ROS/MoveIt and employed in the real robot application.

Table 1 shows the average simulated elapsed time for each manipulator motion performing the loading and the unloading of two boxes placed in the first two columns. The partial time taken to load and unload the first box are indicated as $^sT_{l1}$ and $^sT_{u1}$, respectively, while ^sT_l and ^sT_u represent the overall time needed for loading/unloading two boxes. The left superscript s indicates the shovel-shaped tool. Note also that in Table 1, motion indices are referred to the steps of the first simulated scenario (Procedure 1).

In order to guarantee proper safety bounds, the manipulator joint velocities and the end-effector linear velocity are limited to 40 deg/s and 200 mm/s, respectively.

Table 1. Simulated elapsed time for loading/unloading two aligned boxes.

		Loading	Unloading
	Motion	Time (s)	Time (s)
	1	5.75	5.55
Box	2	1.04	1.04
1	3	3.10	3.05
	4	1.55	1.55
	Partial	$^sT_{l1} = 11.44\,\text{s}$	$^sT_{u1} = 11.19\,\text{s}$
	1	5.25	3.55
Box	2	1.04	1.04
2	3	3.10	3.05
	4	1.55	1.6
	Total	$^sT_l = 22.38\,\text{s}$	$^sT_u = 20.43\,\text{s}$

Therefore, under the further following assumptions:

- T_j is the time required to move the AMR from the initial pallet to the storage pallet;
- $T_s = 1\,\text{s}$ is the time required to move the AMR from the current to the next column of the grid,

and the overall time needed to complete the transport of all the 21 boxes can be fairly approximated as follows:

$$^sT_{\text{tot}} = 22\,T_j + 7\left(^sT_l +^s T_u +^s T_{l1} +^s T_{u1}\right) + 6\,T_s \approx 464 + 22\,T_j \tag{1}$$

where the term $22\,T_j$ considers 11 round-trip journeys, the terms $7\,^sT_l$ and $7\,^sT_u$ include the loading/unloading of the aligned boxes in the first and second grid columns, while $7\,^sT_{l1}$ and $7\,^sT_{u1}$ are the times taken to handle the remaining single box of the third grid column.

It is worth noting that the additional time required to re-execute the procedure in the case of a collision between the tool and the boxes due to the inaccurate detection of the gap, as described in Section 6, is not considered here, because it is expected that the probability of this type of events will be reduced and ideally made negligible for the future industrial release of the system.

The second manipulation scheme (Procedure 2) involves the usage of a vacuum gripper mounted on the robot end-effector instead of the shovel-shaped tool. The main difference with the previous approach is that the robot can execute a standard pick-and-place operation with approaching movements at different heights, so that the lifting device is no more required. Moreover, the detection of the gap between boxes is no longer required and the vision system can identify the picking frame exploiting a single camera acquisition, so that the first robot motion is faster with respect to the former manipulation approach. The algorithm of Procedure 2 describes the sequence of robot motions required to pick a single box from the pallet and place it on the AMR.

Procedure 2: one-box picking/placing, gripper tool

---

1. Point-to-point motion to box target frame
2. Linear downward approaching movement (200 mm)
3. Close the gripper
4. Linear upward approaching movement (200 mm)
5. Linear movement (600 mm)
6. Linear downward approaching movement (200 mm)
7. Open the gripper
8. Linear upward movement (300 mm)

In this case, the loading/unloading procedures are slower than those of the former approach because of the necessity of two additional linear motions during the manipulation phase. The time taken to perform each movement is reported in Table 2. The overall time to complete the transport of the 21 boxes can be computed by means of Equation (1), with the following result:

$$^vT_{\text{tot}} = 22\,T_j + 7\left(^vT_l +\, ^vT_u +\, ^vT_{l1} +\, ^vT_{u1}\right) + 6\,T_s \approx 524 + 22\,T_j \qquad (2)$$

Therefore, the former manipulation scheme provides an overall time saving of $^vT_{\text{tot}} -\,^sT_{\text{tot}} \approx 60\,\text{s}$. Such a result justifies the choice of equipping the AMR with the lifting device and using a tool that allows the manipulator to drag the items instead of picking and placing them.

5. Experimental Setup, Perception and Control System

In order to perform a preliminary evaluation of the proposed depalletizing system, a prototype was set up, as shown in Figure 5a, which does not include the lifting device on the AMR. Moreover, only a single layer of boxes is present. Experiments were conducted in a laboratory environment instead of an industrial setting, due to the ongoing the COVID-19 pandemic. Since we are interested in demonstrating the capability of the system to collect a limited number of items without manipulating an entire pallet layer, we present the experiments that involve unloading of two boxes. The 3D camera is shown in Figure 5b, as well as the paddle tool connected to the manipulator through a custom flange.

Table 2. Simulated elapsed time for loading/unloading two boxes with the vacuum gripper.

		Loading	**Unloading**
	Motion	Time (s)	Time (s)
Box 1	1	4.25	5.90
	2	1.05	1.05
	4	1.05	1.05
	5	3.09	3.10
	7	1.05	1.05
	8	1.55	1.54
	Partial	$^{v}T_{l1} = 12.04\,\mathrm{s}$	$^{v}T_{u1} = 13.69\,\mathrm{s}$
Box 2	1	3.65	3.35
	2	1.05	1.05
	4	1.05	1.00
	5	3.10	3.10
	7	1.05	1.10
	8	1.54	1.54
	Total	$^{v}T_{l} = 23.48\,\mathrm{s}$	$^{v}T_{u} = 24.88\,\mathrm{s}$

The algorithm describing the box extraction task is detailed in Algorithm 1. After the mobile base has moved closer to the pallet, the system performs *detection of parcel boxes* as described in Section 5.1, by moving the 3D camera to an elevated pose in order to have a complete view of the pallet top layer, and it detects the 3D position of all the boxes in front of the loading surface of the robot. Then, in the *box depalletizing plan* phase, the sequence of picking operations to be performed is determined according to the pose of the boxes with respect to the robot. Each picking operation is performed as follows. First, an edge of the box is chosen for the picking and its position is evaluated according to the camera estimation and the box dimensions. Then, the camera on the arm is moved above the estimated position to perform a close view refined estimation of the center of the gap between two boxes, where the tool should be inserted to complete the retrieval (Section 5.2). Once the estimation is obtained, the robot tool is aligned with the gap center and a tool insertion operation is attempted. During the insertion, the tool wrench is continuously evaluated to identify any unexpected collision, possibly due to wrong detection. An admittance control scheme allows to prevent any damages to the robot and the boxes during the insertion as reported in Section 5.3. As soon as a collision occurs, the insertion operation is interrupted, and the estimation of the gap is tried again. If no collision is detected, the insertion is considered successful and the box, dragged by the robot tool, is loaded on the robot support surface, which will be replaced by the lifting device in our future work.
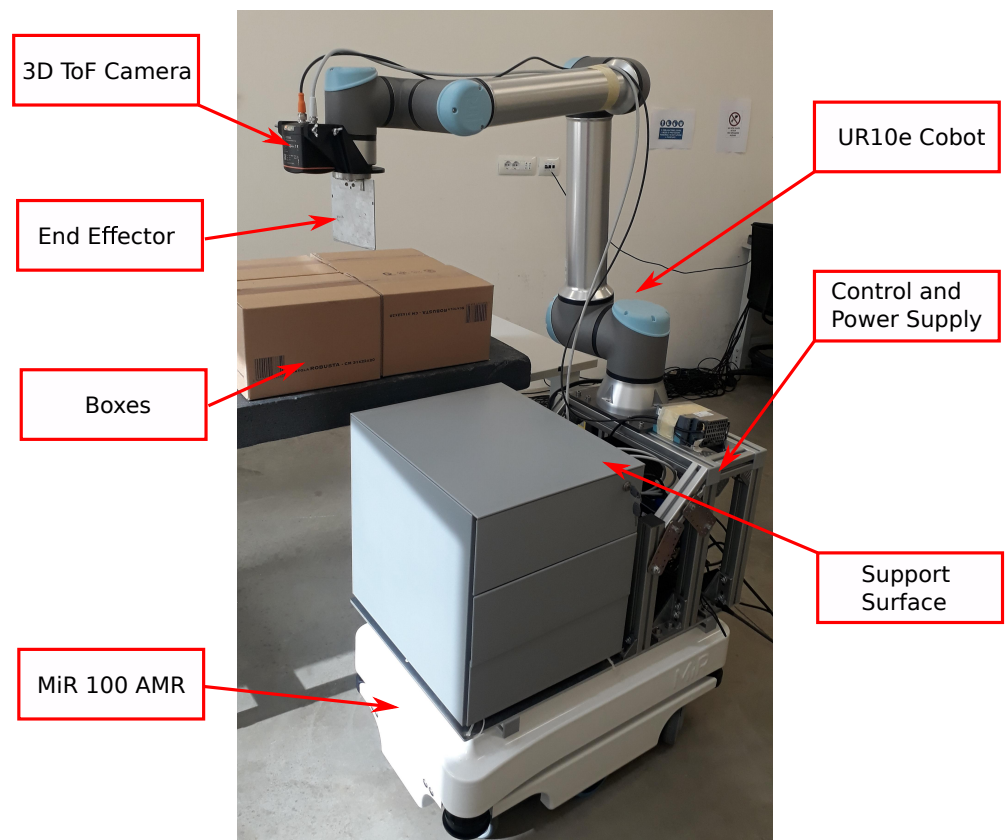
(**a**) Overview of the experimental setup.

(**b**) Detailed view of the end-effector and 3D camera.

Figure 5. The experimental setup: the mobile manipulator (**a**), the paddle tool and the in-hand 3D camera (**b**).

Algorithm 1: Robotized Depalletizing Algorithm

Data: 3D camera image,
  robot tool wrench
Result: depalletizing of a box layer
move camera in the detection pose;
perform *detection of parcel boxes*;
define *box picking sequence* in *box depalletizing plan*;
while *box picking sequence is not empty* **do**
  extraction start: consider the first box in the sequence;
  evaluate the box edge position;
  move the 3D camera to the evaluated position;
  perform *refined estimation of the gap* between boxes using the 3D camera image;
  move the robot tool above the refined gap;
  while *insert the robot tool in the gap* **do**
    if *vertical tool wrench > wrench threshold* **then**
      extract robot tool;
      go to beginning of current section: *extraction start*;
    end
  end
  drag the box toward the robot;
  remove current box from the sequence;
end

5.1. Detection of Parcel Boxes

In the *detection of parcel boxes* phase, the robot arm first moves to an observation configuration, where the upper layer of the pallet can be fully observed by the eye-in-hand infra-red 3D camera (IFM Electronics O3D303). Furthermore, the camera exposure time $t_{\exp}$ is set to a constant value $t_{\exp,\text{far}}$, suitable to observe objects from the current distance of the sensor to the pallet, as shown in Figure 6a. Then, the camera acquires a depth image with a of resolution 352×264 pixels, which is converted into an organized point cloud containing 3D points p_{ij}. The camera also produces an intensity value r_{ij} for each pixel that contains the amount of light returned to the sensor.

As light intensity r_{ij} decreases with distance according to the inverse square law, and since the amount of light received by the sensor is proportional to the current exposure time $t_{\exp}$, a corrected intensity image $R' = \left\{ r'_{ij} \right\}$ is computed as $r'_{ij} = r_{ij} \, \|p_{ij}\|^2 / t_{\exp}$. As the camera provides its own infra-red lighting, both depth and intensity measurements proved robust to changes to environmental light conditions compatible with indoor industrial environments.

In order to ensure stability, pallets are generally arranged as a stack of layers. Each layer above the first one lies on the top planar faces of the parcel boxes of the layer below. In industrial environments, the content and the configuration of a pallet is known in advance. Pallets stored in a warehouse usually contain parcel boxes of the same type. Moreover, palletizing and depalletizing tasks of single parcels are performed by inserting or removing the parcel boxes from the top layer.
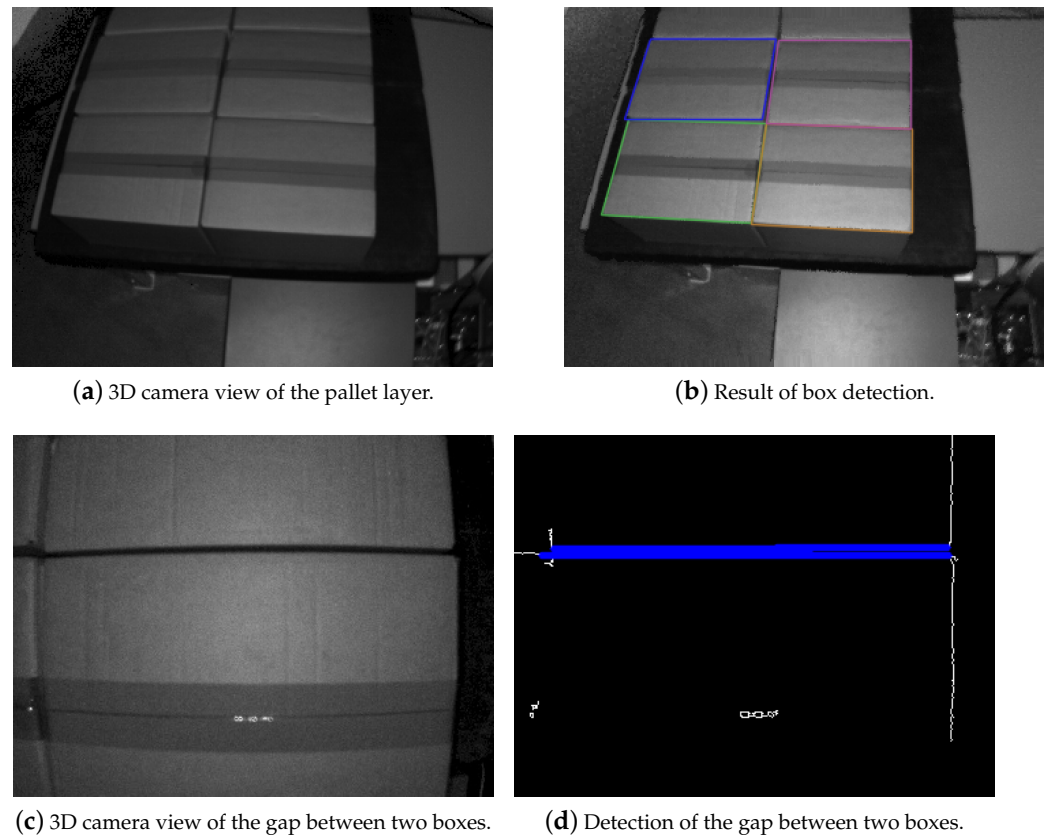
(**a**) 3D camera view of the pallet layer.

(**b**) Result of box detection.

(**c**) 3D camera view of the gap between two boxes.

(**d**) Detection of the gap between two boxes.

Figure 6. Detection of parcel boxes and their separation gaps.

These assumptions derived from industrial practice are exploited to strengthen the robustness of our box detection algorithm.

Hence, in this work, we consider a single type of boxes of known size, but the proposed approach can be easily extended to handle different box formats. Moreover, we can estimate the equation of the top plane of the highest layer of parcel boxes by applying a RANSAC-based estimation method when the pallet layer is full, i.e., when no box has been removed yet. Hence, we assume that the top plane is always known with respect to the robot. Parcel box detection is constrained to the points p_{ij} which are within a small distance from the plane. Parcel boxes are then detected according to the following steps [29]:

5.1.1. Edge Detection and Candidate Boxes Computation

Edges are detected in both the intensity and the depth image acquired by the camera. In particular, discontinuities in the intensity image are detected by applying the standard Canny edge detector, with upper and lower thresholds $U_{canny} = 120$ and $L_{canny} = 60$. Conversely, depth discontinuities are defined as the pixels whose corresponding points belong to the top plane of the pallet (with a tolerance $th_{inlier} = 5$ cm) for which there exists at least a pixel in the neighborhood which does not satisfy the same condition. Straight lines are fitted to the union of the two discontinuity images using the Hough Line Transform.

Lines are organized in a connectivity graph by connecting them at intersections. A set of candidate boxes $\mathbb{B}$ is generated by locating cycles of length 4 in the graph, i.e., quadrilaterals in the image. A candidate box is only accepted if the edges approximately intersect at a right angle, with tolerance $\theta_{box} = 10°$. Moreover, the length of the edges must correspond to the expected size of the box top face (with tolerance $\sigma_{box} = 2$ cm). In the case of a mixed pallet, multiple possible box sizes could be acceptable in this step.

5.1.2. Genetic Optimization Algorithm

As the set of candidate boxes $\mathbb{B}$ detected in the previous step may contain many spurious or overlapping boxes, a genetic optimization algorithm is used to locate the best subset of boxes $\mathbb{S}$. The optimization aims at maximizing the area $F(\mathbb{S})$ of the image (in pixels) which is covered by exactly one of the boxes. In particular, the objective function to be optimized is:

$$F(\mathbb{S}) = \sum_{S \in \mathbb{S}} A(S) - \gamma_O \sum_{S \in \mathbb{S}} O(S) - \gamma_I \sum_{S \in \mathbb{S}} I(S) - \gamma_C \sum_{S \in \mathbb{S}} \sum_{S' \in \mathbb{S} S' \neq S} C(S', S) \tag{3}$$

The first term is the summation of the areas $A(S)$ (in pixels) of the candidate box $S \in \mathbb{S}$, and the remaining three terms are penalty terms where $O(S)$ is the number of pixels of S, which do not belong to the top plane of the pallet, $I(S)$ is the difference between the quadrilateral area and the expected area of the top face of a box, and $C(S, S')$ is the overlapping area between box candidates S and S'. Coefficients $\gamma_O = 2$, $\gamma_I = 8$ and $\gamma_C = 2$ are parameters weighting each penalty term.

The genetic optimization algorithm operates on a population of $G_{\text{pop}} = 100$ individuals $\mathbb{S}_i$, each representing a subset of the candidate box $\mathbb{B}$. The population is initialized by selecting random subsets of candidate boxes. We define two mutation operators Mutation1($\mathbb{S}$) and Mutation2($\mathbb{S}$), and a crossover operator Crossover($\mathbb{S}_1, \mathbb{S}_2$). Operator Mutation1($\mathbb{S}$), applied with probability $P_{\text{M1}} = 0.1$, removes up to three random elements from $\mathbb{S}$. Operator Mutation2($\mathbb{S}$), applied with probability $P_{\text{M2}} = 0.1$, removes a random element from S and replaces it with the element in $\mathbb{B}$ which maximizes $F(\mathbb{S})$. Finally, operator Crossover($\mathbb{S}_1, \mathbb{S}_2$), with probability $P_C = 0.7$, generates a new individual $\mathbb{S}'$ by extracting random elements from $\mathbb{S}_1 \cup \mathbb{S}_2$ as long as they increase the objective function $F(\mathbb{S})$. Furthermore, we define a Fill($\mathbb{S}, \mathbb{B}$) operator, which generates a superset of $\mathbb{S}$ by repeatedly adding random candidate boxes from $\mathbb{B}$ until $F(\mathbb{S})$ stops increasing. By applying the Fill operator at the initialization and after each operator, we ensure that each individual is always a greedy local maximum. Optimization ends when the objective function does not decrease for $G_{\text{stall}} = 20$ consecutive frames.

The genetic optimization algorithm was implemented in C++ using the OpenGA library (https://github.com/Arash-codedev/openGA, accessed on 24 June 2021). The output of the box detection can be appreciated in Figure 6b. Further details on the genetic optimization approach can be found in [30].

5.2. Refined Estimation of the Gap between Two Boxes

In the *refined estimation of the gap* phase, the camera on the arm is moved to the estimated position of the gap between two boxes, at a fixed height above the boxes themselves. The camera is oriented so that the image plane is roughly parallel to the top pallet layer. Moreover, the image x axis is oriented as the expected direction of the box edge. The camera exposure t_{exp} is adjusted to a low value $t_{\text{exp,near}}$ to prevent the saturation of the camera sensor. The 3D camera view at this stage can be seen in Figure 6c.

The gap refinement algorithm consists of three steps. First, 2D lines are detected through the probabilistic Hough Transform on the corrected intensity image R'. Then, 2D lines are discarded if their angle with respect to the image x axis is too large or if they are too far from the image center, as shown in Figure 6d. Through a proper choice of thresholds, only the horizontal lines in the center of the image remain and most outliers are discarded. Finally, the gap position p_{ref} and orientation θ_{ref} are computed as the mean position and orientation of the detected lines, and used to determine the target frame for the manipulation operation.

5.3. Controller Design

During the extraction procedure, an admittance control strategy is implemented for the control of the arm position x_{ref}:

$$x_{\text{ref}} = x_{\text{des}} + \Lambda \Delta x_{\text{des}} \tag{4}$$

along the tool insertion direction defined by the selection matrix Λ:

$$\Lambda = [0\ 0\ 1\ 0\ 0\ 0]^T \tag{5}$$

The admittance behavior is based on the tool wrench measure provided by the cobot and it produces a smooth behavior during tool insertion. In particular, a displacement for the control reference is defined at each time instant according to the measured wrench, in order to produce an elastic interaction in the case of collisions and prevent any damage to the boxes. The induced displacement Δx_{des} is obtained by means of a first-order digital filter to reduce the wrench estimation noise:

$$\Delta x_{\text{des}} = \frac{K_F}{1 + K_P\, z^{-1}}\, \text{dz}(\Lambda^T F_{\text{ext}}, F_{\text{th}}) \tag{6}$$

where z^{-1} represents the sample-time delay according to the Z-transform notation and K_p is the filter gain. The wrench signal is evaluated with respect to a predefined deadzone function $\text{dz}(\cdot, \cdot)$ to prevent undesired oscillations of the control:

$$\text{dz}(F_1, F_2) = \begin{cases} F_1 - F_2, & F_1 > F_2 \\ 0, & -F_2 < F_1 < F_2 \\ F_1 + F_2, & F_1 < -F_2 \end{cases} \tag{7}$$

6. Experimental Results

Experiments were performed to show the capabilities of the robot platform and to evaluate its effectiveness in depalletizing tasks. We report unloading tasks in scenarios where the mobile platform is already in place to collect the cardboard boxes, as well as in more dynamic cases where the AMR first approaches the boxes. The experiments were performed with filter gains $K_F = 10e^{-3}$, $K_P = 0.1$ and force threshold $F1 = 10N$.

6.1. 2-Boxes Depalletizing and Transportation

A two-boxes depalletizing task is shown in Figure 7. First, the mobile base approaches a group of boxes by moving to a suitable configuration for the box unloading task, then the cobot rises the 3D camera to an elevated position to detect the boxes. The two-steps detection of the gap between adjacent boxes was then executed. Then, the tool is inserted into the gap between the first two boxes and the first box is pulled on the robot. Then, the procedure is repeated to manipulate the second box. Finally, the mobile base moves away with the two boxes onboard. In Figure 8, the Cartesian position and the measured wrench on the cobot end-effector are displayed along the task. The low values of the measured wrench confirm that the task was successfully completed without any collision of the tool.

Figure 7. From top-left to bottom-right: a sequence showing the execution of a 2-boxes depalletizing task by the mobile manipulator.
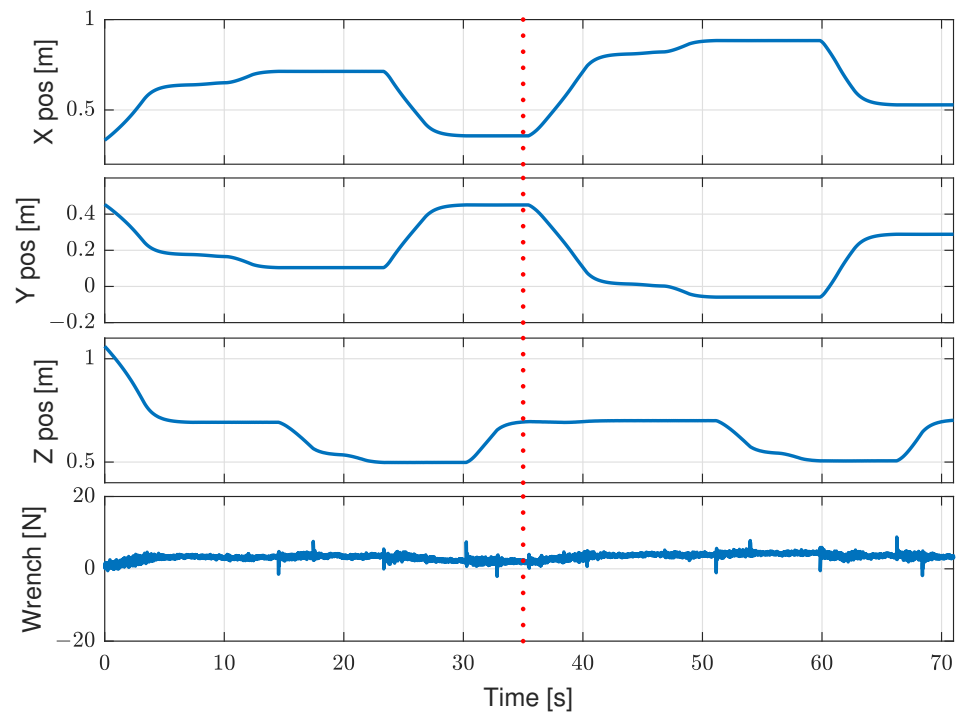
Figure 8. A depalletizing task where two boxes are successfully extracted. The mobile base is exploited to approach the boxes for the unloading task and to transport the boxes away once loaded on the robot.

Figure 9 reports the success rate of the task for 50 box extractions. It can be noticed that in most cases, the system performed the extraction by the first attempt. Nonetheless, in a few cases, the first estimation of the gap between boxes was not correct, and a second

attempt was required. This is probably due to unexpected changes in light conditions or slight motions of the mobile base. In all cases, the second estimation of the gap was successfully completed and the box was correctly loaded onto the robot.
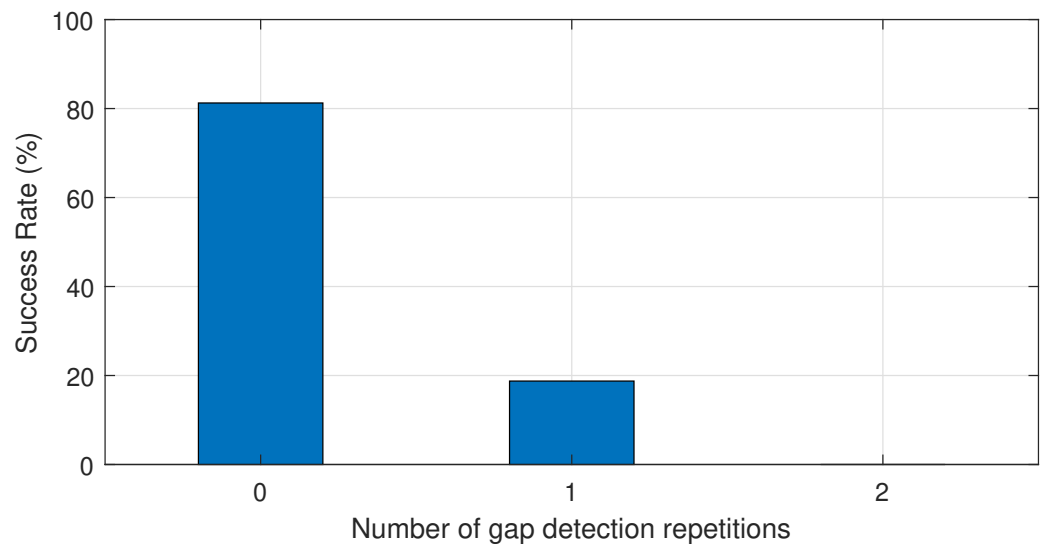

Figure 9. Overall insertion success rate with respect to gap detection repetitions.

6.2. Single Box Extraction

The Cartesian position and the measured wrench on the cobot end-effector during the unloading of a single box are reported in Figure 10a. At first, the box position was estimated by the in-hand camera. Then, the estimated gap between two adjacent boxes was refined to obtain the target frame for tool insertion ($t = 4$ s). Once the target frame was determined, the tool was aligned with the gap between the boxes ($t = 8$ s), and an insertion attempt was performed ($t = 12.5$ s). During the insertion phase, a collision was detected ($t = 15.5$ s), possibly caused by an inaccurate evaluation of the gap position. Therefore, tool insertion was suspended and a new detection of the gap was performed ($t = 21$ s). Since no further collisions were detected ($t = 32$ s) the cobot completed the insertion and the box was successfully pulled onto the robot ($t = 36$ s). It can be noticed that when a collision occurs, the admittance control allows for a smooth motion of the end-effector, thus reducing the risk of damaging the robot arm or the target box.

6.3. Complete Layer Depalletizing

A successful example of unloading a pallet layer with four boxes, in a static scenario where the AMR was fixed, is shown in Figure 10a. In this test case, the target boxes were extracted one by one until the pallet layer was empty. The AMR was placed so that all four boxes were inside the robot arm workspace. The boxes were manually removed from the robot platform once unloaded since the robot platform can carry at most two boxes. The graph in Figure 10b shows that the system was able to successfully extract all four boxes. A collision occurred when the tool was inserted to extract the third box, and a second estimation of the gap was required for the task to be successfully completed. It is noteworthy that each extraction showed a similar behavior, and a comparable completion time (with the exception of the collision case). Moreover, the robot was moving at only 10% its maximum speed. Therefore, the expected performance obtained in Section 4 can be easily achieved if the robot arm was moving at full speed.
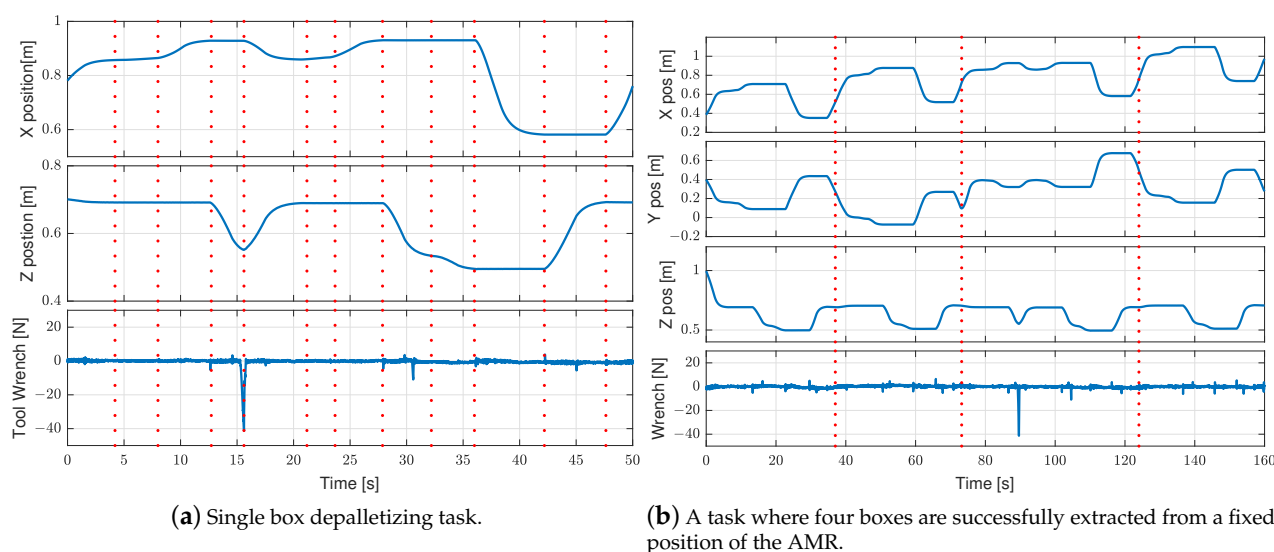
(**a**) Single box depalletizing task.

(**b**) A task where four boxes are successfully extracted from a fixed position of the AMR.

Figure 10. Robot tool position and wrench during a depalletizing task of a single box (**a**) and four boxes (**b**).

7. Conclusions and Future Work

In this paper, a novel mobile manipulator equipped with a 3D perception system for autonomous depalletizing tasks was presented. The proposed solution shows that it is possible to unload, in a controlled way, just a few boxes from a pallet, without the need of disassembling an entire pallet layer. Therefore, the system can be very effective to build mixed pallets containing different types of goods, in which only a small number of boxes of the same type are needed.

Future work will be devoted to the evaluation of the proposed system in an industrial scenario, which was not possible due to COVID-19-related restrictions. Moreover, the implementation of the depalletizing task will be extended to handle packages of bottles and cans, which will require a more complex manipulation plan. We will also consider the task of building a complete mixed pallet, including the collection of small lots of boxes from multiple single-item pallets and their arrangement on the mixed pallet in a proper order.

Author Contributions: Conceptualization, J.A., A.B., M.C., G.P., L.S.; methodology, R.M., M.B., A.B., D.L.R., G.P., F.Z.; software, R.M., D.C., J.R.; validation, G.S., F.Z., D.C., J.R.; writing—original draft preparation, R.M., J.A., F.Z., A.B., R.D.L., G.I., S.F.; writing—review and editing, M.C., L.S.; supervision, J.A., R.D.L., C.F., G.I., M.C., D.L.R., C.M., L.S.; funding acquisition, J.A., C.F. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Echelmeyer, W.; Kirchheim, A.; Wellbrock, E. Robotics-logistics: Challenges for automation of logistic processes. In Proceedings of the 2008 IEEE International Conference on Automation and Logistics, Qingdao, China, 1–3 September 2008; pp. 2099–2103.
2. Khairuddin, U.; Razi, N.; Abidin, M.; Yusof, R. Smart Packing Simulator for 3D Packing Problem Using Genetic Algorithm. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2020; Volume 1447, p. 012041.
3. Al-Jodah, A.A.L.; Shirinzadeh, B.; Pinskier, J.; Ghafarian, M.; Das, T.K.; Tian, Y.; Zhang, D. Antlion Optimized Robust Control Approach for Micropositioning Trajectory Tracking Tasks. *IEEE Access* **2020**, *8*, 220889–220907. [CrossRef]
4. Al-Azza, A.A.; Al-Jodah, A.A.; Harackiewicz, F.J. Spider monkey optimization (SMO): a novel optimization technique in electromagnetics. In Proceedings of the 2016 IEEE Radio and Wireless Symposium (RWS), Austin, TX, USA, 24–27 January 2016; pp. 238–240.
5. Li, Q.; Dong, S.; Zhang, D.; Wang, X. Research on the Lidar-based Recognition and Location Method for Depalletizing Targets. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020; pp. 683–687.

6. Wei, C.; Ji, Z.; Cai, B. Particle swarm optimization for cooperative multi-robot task allocation: A multi-objective approach. *IEEE Robot. Autom. Lett.* **2020**, *5*, 2530–2537. [CrossRef]

7. Sabattini, L.; Aikio, M.; Beinschob, P.; Boehning, M.; Cardarelli, E.; Digani, V.; Krengel, A.; Magnani, M.; Mandici, S.; Oleari, F.; et al. The PAN-Robots Project: Advanced Automated Guided Vehicle Systems for Industrial Logistics. *IEEE Robot. Autom. Mag.* **2018**, *25*, 55–64. doi:10.1109/MRA.2017.2700325. [CrossRef]

8. Cesetti, A.; Scotti, C.; Di Buo, G.; Longhi, S. A service oriented architecture supporting an autonomous mobile robot for industrial applications. In Proceedings of the 18th Mediterranean Conference on Control and Automation, MED'10, Marrakech, Morocco, 23–25 June 2010; pp. 604–609.

9. Pedrosa, E.; Lim, G.H.; Amaral, F.; Pereira, A.; Cunha, B.; Azevedo, J.L.; Dias, P.; Dias, R.; Reis, L.P.; Shafii, N.; et al. TIMAIRIS: Autonomous Blank Feeding for Packaging Machines. In *Bringing Innovative Robotic Technologies from Research Labs to Industrial End-Users*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 153–186.

10. Bonini, T.; Forni, A.; Mazzolini, M. Design of an Intelligent Handling System using a Multi-Objective Optimization Approach. In Proceedings of the IEEE 23rd Internationl Conference on Emerging Technologies and Factory Automation (ETFA), Torino, Italy, 4–7 September 2018; Volume 1, pp. 887–894.

11. Doliotis, P.; McMurrough, C.D.; Criswell, A.; Middleton, M.B.; Rajan, S.T. A 3D perception-based robotic manipulation system for automated truck unloading. In Proceedings of the IEEE Internationl Conference on Automation Science and Engineering (CASE), Fort Worth, TX, USA, 21–25 August 2016; pp. 262–267.

12. Caccavale, R.; Arpenti, P.; Paduano, G.; Fontanellli, A.; Lippiello, V.; Villani, L.; Siciliano, B. A Flexible Robotic Depalletizing System for Supermarket Logistics. *IEEE Robot. Autom. Lett.* **2020**, *5*, 4471–4476. [CrossRef]

13. Nakamoto, H.; Eto, H.; Sonoura, T.; Tanaka, J.; Ogawa, A. High-speed and compact depalletizing robot capable of handling packages stacked complicatedly. In Proceedings of the 2016 IEEE/RSJ Internationl Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 344–349.

14. Kavoussanos, M.; Pouliezos, A. Visionary automation of sack handling and emptying. *IEEE Robot. Autom. Mag.* **2000**, *7*, 44–49. [CrossRef]

15. Matsuo, I.; Shimizu, T.; Nakai, Y.; Kakimoto, M.; Sawasaki, Y.; Mori, Y.; Sugano, T.; Ikemoto, S.; Miyamoto, T. Q-bot: Heavy object carriage robot for in-house logistics based on universal vacuum gripper. *Adv. Robot.* **2020**, *34*, 173–188. [CrossRef]

16. Lim, G.H.; Pedrosa, E.; Amaral, F.; Lau, N.; Pereira, A.; Azevedo, J.L.; Cunha, B.; Badini, S. Mobile manipulation for autonomous packaging in realistic environments: EuRoC challenge 2, stage II, showcase. In Proceedings of the 2018 IEEE Internationl Conference on Autonomous Robot Systems and Competitions (ICARSC), Torres Vedras, Portugal, 25–27 April 2018; pp. 231–236.

17. Dogar, M.R.; Srinivasa, S.S. A planning framework for non-prehensile manipulation under clutter and uncertainty. *Auton. Robot.* **2012**, *33*, 217–236. [CrossRef]

18. Papallas, R.; Dogar, M.R. Non-prehensile manipulation in clutter with human-in-the-loop. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Virtual Workshops, 31 May–30 June 2020; pp. 6723–6729.

19. Acharya, P.; Nguyen, K.D.; La, H.M.; Liu, D.; Chen, I.M. Nonprehensile Manipulation: a Trajectory-Planning Perspective. *IEEE/ASME Trans. Mechatron.* **2020**, *26*, 527–538. [CrossRef]

20. Ardakani, M.; Bimbo, J.; Prattichizzo, D. Quasi-static Analysis of Planar Sliding Using Friction Patches. *arXiv* **2019**, arXiv:1904.06677.

21. Qin, Y.; Escande, A.; Tanguy, A.; Yoshida, E. Vision-based Belt Manipulation by Humanoid Robot. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Online Metting, 25–29 October 2020; pp. 3547–3552.

22. Hashimoto, M.; Sumi, K. Genetic labeling and its application to depalletizing robot vision. In Proceedings of the 1994 IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 5–7 December 1994; pp. 177–186.

23. Yunardi, R.T.; Winarno, P. Contour-based object detection in Automatic Sorting System for a parcel boxes. In Proceedings of the 2015 International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation (ICAMIMIA), Surabaya, Indonesia, 15–17 October 2015; pp. 38–41.

24. Prasse, C.; Stenzel, J.; Böckenkamp, A.; Rudak, B.; Lorenz, K.; Weichert, F.; Müller, H.; ten Hompel, M. New Approaches for Singularization in Logistic Applications Using Low Cost 3D Sensors. In *Sensing Technology: Current Status and Future Trends IV*; Mason, A., Mukhopadhyay, S.C., Jayasundera, K.P., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 191–215.

25. Katsoulas, D.; Bastidas, C.C.; Kosmopoulos, D. Superquadric Segmentation in Range Images via Fusion of Region and Boundary Information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 781–795. [CrossRef] [PubMed]

26. Zhang, B.; Skaar, S.B. Robotic de-palletizing using uncalibrated vision and 3D laser-assisted image analysis. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, USA, 11–15 October 2009; pp. 3820–3825.

27. Arpenti, P.; Caccavale, R.; Paduano, G.; Andrea Fontanelli, G.; Lippiello, V.; Villani, L.; Siciliano, B. RGB-D Recognition and Localization of Cases for Robotic Depalletizing in Supermarkets. *IEEE Robot. Autom. Lett.* **2020**, *5*, 6233–6238. [CrossRef]

28. Baldassarri, A.; Innero, G.; Di Leva, R.; Palli, G.; Carricato, M. Development of a Mobile Robotized System for Palletizing Applications. In Proceedings of the 25th IEEE International Conference on Emerging Technologies and Factory Automation, Vienna, Austria, 8–11 September 2020; Volume 1, pp. 395–401.

29. Chiaravalli, D.; Palli, G.; Monica, R.; Lodi Rizzini, D.; Aleotti, J. Integration of a Multi-Camera Vision System and Admittance Control for Robotic Industrial Depalletizing. In Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria, 8–11 September 2020; Volume 1, pp. 667–674.

30. Monica, R.; Aleotti, J.; Rizzini, D.L. Detection of Parcel Boxes for Pallet Unloading Using a 3D Time-of-Flight Industrial Sensor. In Proceedings of the 2020 Fourth IEEE International Conference on Robotic Computing (IRC), Virtual Conference, 9–11 November 2020; pp. 314–318.