A systematic strategy of pallet identification and picking based on deep learning techniques

Article *in* Industrial Robot the international journal of robotics research and application · January 2023

DOI: 10.1108/IR-05-2022-0123

6 authors, including:

Yongyao Li
Changchun University of Science and Technology
3 PUBLICATIONS **19** CITATIONS
SEE PROFILE

Qinglei Zhao
Changchun Institute of Optics, Fine Mechanics and Physics
11 PUBLICATIONS **137** CITATIONS
SEE PROFILE

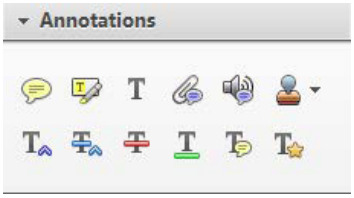
Qi Song
Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sci…
30 PUBLICATIONS **178** CITATIONS
SEE PROFILE

Smart Proof System Instructions

It is recommended that you read all instructions below; even if you are familiar with online review practices.

Using the Smart Proof system, proof reviewers can easily review the PDF proof, annotate corrections, respond to queries directly from the locally saved PDF proof, all of which are automatically submitted directly to **our database** without having to upload the annotated PDF.

- ✓ **Login into Smart Proof** anywhere you are connected to the internet.
- ✓ **Review the proof** on the following pages and mark corrections, changes, and query responses using the **Annotation Tools**.



Note: Editing done by replacing the text on this PDF is not permitted with this application.

- ✓ **Save your proof corrections** by clicking the "Publish Comments" button.
  Corrections don't have to be marked in one sitting. You can publish comments and log back in at a later time to add and publish more comments before you click the "Complete Proof Review" button below.

- ✓ **Complete your review** after all corrections have been published to the server by clicking the "Complete Proof Review" button below.

Before completing your review.....

Did you reply to all author queries found in your proof?

Did you click the "Publish Comments" button to save all your corrections? Any unpublished comments will be lost.

Note: Once you click "Complete Proof Review" you will not be able to add or publish additional corrections.

AUTHOR QUERIES

Note: *It is crucial that you NOT make direct edits to the PDF using the editing tools as doing so could lead us to overlook your desired changes. Edits should be made via the 'comments' feature.*

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1— Please confirm the given-names and surnames are identified properly by the colours.
■=Given-Name, ■= Surname
The colours are for proofing purposes only. The colours will not appear online or in print.

AQ2— Please check the accuracy of the affiliation(s) of each author and make changes as appropriate. Affiliations cannot be changed once the article has been published online. Please ensure to include the city and country names in the affiliation(s), as these are mandatory in line with Emerald house style.

AQ3— Please check the correctness of the affiliations and amend as and if necessary.

AQ4— Please note the addition of "Department of" to the affiliation "Changchun University of Science and Technology, Jilin, China", "Pilot AI Company, Hanghzou, China", "Changchun University of Science and Technology, Jilin, China" and "Suzhou Institute of Biomedical Engineering and Technology, Suzhou, China". We request you to verify the accuracy and appropriateness of this.

AQ5— Please note that to conform to the journal guidelines (i.e. using the phrase "the purpose of this study/paper. . ." or "this study/paper aims to. . ." in Purpose section of the Abstract), we have revised the following sentence in the Purpose of the Abstract. Please check and amend if necessary: "This paper aims to present a novel approach . . .".

AQ6— Please provide the spelled-out forms of the following abbreviations: RGB, LiDAR, VGG, if necessary. Ignore if standard abbreviation.

AQ7— You have used "data" in the singular form in the text and we have retained your intended meaning. However, if you wish to imply its plural context, revisions with respect to its associated verb usage will need to be made. Please check and amend if necessary.

AQ8— Please check the following sentence for clarity and amend as necessary: "The pallet's Point Cloud data is correlated with . . .".

AQ9— Note that both the terms "Single Shot Detector" and "Single Shot Multi-box Detector" are abbreviated as "SSD" in the paper. Please check the terms for correctness and revise if necessary.

AQ10— Please note that we have considered spell-out form of abbreviation "RGB-D" as "RGB-Depth". Please check and amend if necessary.

AQ11— Please check the following sentence for clarity and amend as necessary: "For the Faster R-CNN, in its first . . .".

AQ12— There are currently no acknowledgements included. Please confirm if this is correct or provide the acknowledgements.

AQ13— Please provide complete author details for all et al. references.

AQ14— Please provide first column head for Tables 5, 9 and 10.

A systematic strategy of pallet identification and picking based on deep learning techniques

AQ:1 Yongyao Li
AQ:2 Department of, Changchun University of Science and Technology, Jilin, China
Guanyu Ding and Chao Li
Department of, Pilot AI Company, Hanghzou, China
Sen Wang
AQ: 3 Department of, Changchun University of Science and Technology, Jilin, China
Qinglei Zhao
AQ: 4 Changchun Institute of Optics Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, and
Qi Song
Department of, Suzhou Institute of Biomedical Engineering and Technology, Suzhou, China

Abstract

AQ: 5 **Purpose** – This paper aims to present a novel approach of pallet identification and localization algorithm (PILA) based on RGB image and depth
AQ: 6 data and the vehicle control algorithm to align with the identified pallet.

Design/methodology/approach – The algorithm pipeline is proposed for real-time running, and the RGB and depth data from the low-cost RGB-D camera. Deep neural network method is applied to detect and locate the pallet in the RGB images. The pallet's point cloud data is correlated with the labelled region of interest in RGB images and the pallet's front-face plane is extracted from point cloud data and the precise orientation of the pallet is
AQ: 7 obtained.The triangle-centric points of pallet's front-face are determined by forming T-shape as a general geometrical rule. The vehicle alignment
AQ: 8 algorithm is proposed to implement the vehicle approaching and precise pallet picking operation as "proof-of-concept" to test PILA performance.

Findings – Experimentally, the orientation angle and centric location of the two kinds of pallets are investigated without any artificial marking. The results show that the pallet could be located with a three-dimensional localization accuracy of 1 cm and an angle resolution of 0.4 degrees at a distance of 3 m with the vehicle control algorithm.

Research limitations/implications – PILA's performance is limited by the current depth camera's range (< = 3 m), and this is expected to be improved by using a better depth measurement device in the future.

Originality/value – The results prove that PILA can work with a real warehouse robot system to deliver the precise pallet location and is applicable for autonomous pallet-picking instruments and self-driving forklift applications.

Keywords Deep neutral network, Pallet localization vehicle control, Pallet recognition, RGBD camera

Paper type Technical paper

1. Introduction

Recently forklift robots have received broad attention in logistics applications, as during the COVID-19 pandemic all over the world, fully automated logistics pallet picking using an unmanned forklift or automated guided vehicle (AGV) became an urgent desire (Song *et al.*, 2020). The massive goods deployment requires the following critical issues to be addressed:

- Forklift robots must know the precise x, y and z coordinates and the angle of the pallet surface centre.
- "Real-time" requirement must be met, and fast computing is necessary to guarantee the pallet auto-picking efficiency.

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/0143-991X.htm**

Industrial Robot: the international journal of robotics research and application

[DOI 10.1108/IR-05-2022-0123]

Funding: This work was financially supported by the National Natural Science Foundation of China (#61975228) and Dalian Technology Bureau (Covid-19 Emergency Fund) and Jilin Science and Technology Development Plan Project (Grant # 20220203053SF).

Conflicts of interest/competing interests: The authors declare no conflict of interest.

Code or data availability: TDS software demo using PILA can be found at the following link or contact: qsong@soleilwares.com.
Source code: www.github.com/unlogical0327/TDS_V1/tree/master/src.
www.github.com/ChinaLyy/VAA.
Pallet Data set: www.github.com/unlogical0327/Pallet_database.

Author contributions: Conceptualization, Gunayu Ding and Qi Song; data curation, Yongao Li; formal analysis, Yongyao Li, Qi Song and Guanyu Ding; investigation, Chao Li, Qinglei Zhao and Sen Wang; methodology, Qi Song, Guanyu Ding, Chao Li and Yongyao Li; software, Qi Song, Guanyu Ding and Chao Li; supervision, Qi Song, validation, Yongyao Li and Qi Song; writing – original draft, Yongyao Li and Qi Song; writing – review and editing, Yongyao Li, Qinglei Zhao and Qi Song; All authors have read and agreed to the published version of the manuscript.

Ethics approval: Not applicable.

Consent to participate: Not applicable.

Consent for publication: Agree.

Received 9 May 2022
Revised 18 June 2022
16 September 2022
22 November 2022
Accepted 23 November 2022

J_ID: IR ART NO: 10.1108/IR-05-2022-0123 Date: 28-December-22 Page: 2 Total Pages: 13 4/Color Figure(s) ARTTYPE="ResearchArticle"

- The pallet types and sizes may be randomly presented in practice, whereas the typical model-based pallet localization methods cannot handle all the cases properly.

To solve the above issues, vision-based and point cloud data solutions have already been proposed to provide automatic pallet localization. However, both solutions use a single data source, i.e. RGB imagery or point cloud data, and are limited by either localization precision or recognition accuracy. For example, researchers used a two-dimensional (2D) laser rangefinder and applied feature detection to locate the pallets (Wang *et al.*, 2021; Tsiogas *et al.*, 2021). However, this method suffered from pallets varying in size and shape, as it could not capture enough features from the 2D depth information (Mohamed *et al.*, 2020; Molter and Fottner, 2018). Some authors suggested using plane segmentation and template matching or registration to deliver more precise results (Xiao *et al.*, 2017). However, the speed and accuracy of pallet recognition are heavily influenced by pallet types and the quality of point cloud data, potentially imposing severe requirements requiring a demanding high-fidelity depth camera and a powerful computer, significantly increasing cost. Currently, all existing methods use a single data source like RGB images or point cloud and suffer from a high probability of false positioning or consume substantial computing power and raise the cost dramatically (Fontana *et al.*, 2021; Mok *et al.*, 2021; Syrjänen, 2021; Ward *et al.*, 2021). Thus, this work presents a novel systematic strategy to identify and pick high-precision pallets. Our core algorithm is designed based on the pallet identification and localization algorithm (PILA) using a vehicle alignment algorithm (VAA) with low-cost hardware. In the first part of this strategy, a deep neural network (DNN) (Srinidhi *et al.*, 2021) is used to detect and locate pallets in RGB images. Then, the point cloud data of the pallet are correlated with the region of interest (RoI) from the RGB image, and after that, the pallet's central-point locations and front surface angle are calculated through plane-fitting and geometric feature extraction using the "T-shape" rule. Finally, the pallet location and angle are sent to VAA to align the forklift's arm with the pallet pockets. The developed algorithm is implemented in C++ for running efficiency, and a controller area network (CAN) protocol is chosen to communicate with the AGV (Jiao *et al.*, 2019). The experimental results demonstrate that the recognition rate is about 1.4 per second, the accuracy is 1 cm and the angular error is below 0.4 degrees at a distance of 3 m.

The remainder of this paper is organized as follows. Sections 1 and 2 introduce and review the related work. Section 3 presents the proposed PILA structure, the detailed RGB training processes using the Single Shot Detector (SSD) and the point cloud processing. Section 4 describes the VAA algorithm flowchart, and Section 5 provides the experimental results of PILA and the simulation results using VAA. Finally, Section 6 summarizes and concludes this work and provides future research directions. AQ: 9

2. Related work

The research on pallet localization can be divided into two categories: vision-based using a 2D camera [Figure 1(a)] and ranging data-based using 2D laser scanners or 3-dimensional (3D) cameras, which can acquire precise depth information in the form of 2D or 3D data [Figure 1(b)]. F1

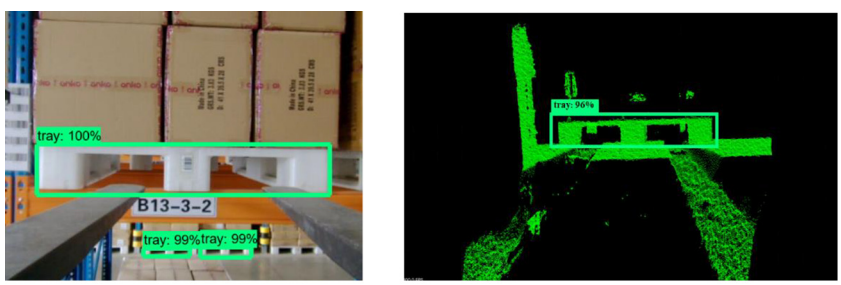
2.1 Vison-based systems

Such systems usually require artificially added features on the pallet, like fiducial markers (Tsiogas *et al.*, 2021), or using the pallet's features, such as colour, shape and other structural information. However, associating with the pallet is highly undesirable in the daily warehouse operation. For example (Iinuma *et al.*, 2020) used computer vision techniques to detect the pallet's two central pockets and then estimated their geometric centres. However, the dimensions and geometry of the pallets must be known a priori, and this method is highly parameter-dependent. Given the wide variety of pallets in a real warehouse environment, this method is somehow impractical. An accurate image segmentation method based on colour and geometric characteristics has also been proposed to detect the pallet's central points (Deshpande *et al.*, 2020; Jia *et al.*, 2021a). However, this approach requires stringent lighting and camera calibration. A deep learning framework has recently been proposed to detect pallets from 2D images (Li *et al.*, 2019). This method has demonstrated a strong potential to handle various pallet types and affords a more accurate identification rate and localization, but it also suffers from poor lighting or defects.

2.2 Range data-based systems

These methods use ranging sensors, e.g. 2D laser rangefinder or 3D LiDAR, which deliver the ranging data in centimetres or even millimetres. Such a method requires more computing

Figure 1 (a) Vision-based pallet localization and (b) depth camera used on forklift to localize pallet



(a) (b)

J_ID: IR ART NO: 10.1108/IR-05-2022-0123 Date: 28-December-22 Page: 3 Total Pages: 13 4/Color Figure(s) ARTTYPE="ResearchArticle"

power and leads to a high cost for real-time applications (Molter and Fottner, 2018). For instance, a 2D laser scanner has been used to detect the artificial reflectors mounted on a target pallet in advance. In (Walter *et al.*, 2010), a closed-loop pallet manipulation system was constructed, which is an estimation framework that uses a sequence of classifiers to infer the structure and posture of objects from individual LiDAR scans. In both works, pallet positioning required detecting the nearest pallet edges at the front face from a single 2D laser scan. Similarly, 3D laser scanner data has been used to detect, locate and track pallets through a machine-learning approach (Mohamed *et al.*, 2020). In this case, the 3D data was converted to bitmap pictures, which were then input to a CNN to search for potential pallets. Another deep learning methodology was proposed using a time-of-flight camera to improve accuracy in warehouse situations. This method was based on point cloud plane contour matching (Wenhan *et al.*, 2019; Jia *et al.*, 2021b). However, the numerous pallet types are difficult to recognize without colour information.

To be deployed in the logistics industry, a pallet picking system must handle more complex cases, such as occlusions by random objects, incomplete surfaces or different lightning. T1 Table 1 highlights that a hybrid system can achieve high recognition rates based on vision-based methods and accurate localization from range-based methods.

In this work, the proposed method achieves a higher recognition rate and localization accuracy at a reasonably rapid detection rate on a computer with an Intel sixth-T2 generation CPU. Table 2 reports that our PILA approach outperforms the previously published studies. Moreover, we also create a VAA to control the forklift to pick up the detected pallets successfully. Our solution is accurate, cost-effective and can contribute to the logistics industry's automation needs.

3. Model description

This section introduces the PILA pipeline strategy in more F2 detail. Figure 2 illustrates the RGB and depth images provided by an RGB-Depth (RGB-D) camera. Then, a DNN is used to AQ: 10 recognize the possible pallet from the RGB images in various scenes, and the model is generated by offline training it on the pallet data set. Then, the well-trained learning model is used for online pallet detection using images from real-time camera readings. The developed algorithm is divided into three stages, as illustrated in Figure 2. In the first stage, the pallet is detected, and the confidence score is passed to the next step. In the second stage, the RGB-D images are used to align the pallet in the RGB image with the corresponding depth image. In the third stage, the cropped point cloud data is extracted for the pallet front-face plane, and then line segments are picked to locate a "T-shape" of the pallet centre. In particular, the horizontal (x) and vertical (y) line segments at the pallet's edge are detected according to the pallet shape, which may vary across the pallet types. The decision rule used here is designed to find the "T-shape" of the pallet centre as a universal solution compensating for the various pallet sizes. The intersection points of the x- and y-line segments define the middle section of the pallet, whereas three points, A, B and C, between the two pockets of the pallet, are used to locate the pockets and centre (Figure 3). Meanwhile, the x, y and z values of the pallet's F3 central location and the angle of the pallet facet are calculated.

3.1 Neural network training for pallet recognition

Object detection architectures are typically classified as one-stage and two-stage detectors. The former commonly includes the Single Shot Multi-box Detector (SSD) and You Only Look Once (YOLO). The latter type involves the regional proposed network (RPN) method, reginal convolutional neuron network (R-CNN) and Faster R-CNN (Zou *et al.*, 2019).

For a single-stage detector like YOLO, the input image is divided into a few grid cells that predict a fixed number of bounding boxes with an intersection over the union (IoU) score. The output is obtained by multiplying the object detection probability with IoU, which is the overlap ratio between the area of the predicted bounding box and the ground-truth bounding box (Redmon *et al.*, 2016).

Table 1 Comparison of range data-based system and vision-based systems

Method	Range data-based system	Vison-based system	Hybrid system
Information source	Depth data	Patterns, colours, texture	Depth data, pattern, colours, texture
Pattern recognition algorithm	Hard to implement	Well developed	Easy to implement
Recognition rate	Slow	Fast	Medium
Recognition accuracy	Low	High	High
Localization accuracy	Very high and reliable	High with false results	Very high and reliable
Computing complexity	High	Low	Medium

Table 2 Comparison of commercial solutions with PILA

Method	Depth image (Molter and Fottner, 2018)	Range and look pallet finder (RLPF) (Behrje *et al.*, 2018)	2D images (Casado *et al.*, 2017)	PILA
Speed (ms)	900	NA	approximately 50	700
Accuracy (mm/degree)	30/NA	10.8/0.78	15/0.57	9.9/0.4
Distance (m)	3.5	4	4.5	3

J_ID: IR ART NO: 10.1108/IR-05-2022-0123 Date: 28-December-22 Page: 4 Total Pages: 13 4/Color Figure(s) ARTTYPE="ResearchArticle"

Figure 2 Flowchart of PILA, which consists of DNN trained module, Online detection module, Fusion module, Point Cloud process module and T-shape forming module

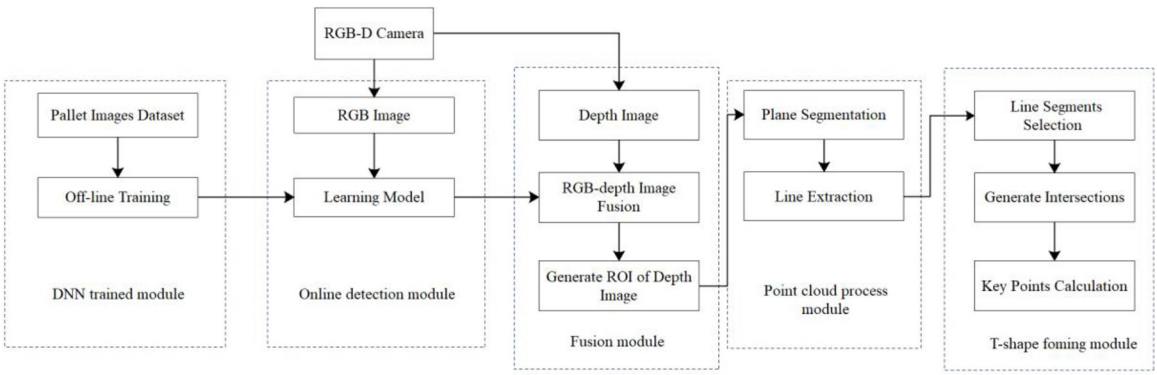
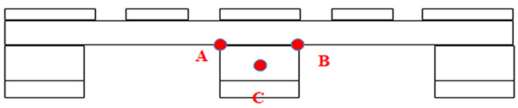


Figure 3 Triangle points of pallet's front surface view



The SSD method uses a whole image as input and passes it over multiple convolution layers to create the corresponding convolutional feature maps, which are used to predict the bounding boxes. The model generates a vector of object class probabilities for predicting bounding boxes. The SSD architecture is illustrated in Figure 4, which is a VGG-16 model pre-trained on ImageNet for image classification. A feed-forward convolutional network generates a fixed-size set of bounding boxes and the classification score. For the SSD model, instead of predicting a possible object score, it provides the likelihood of a class being present in the bounding box.

Faster R-CNN is the most popular two-stage architecture that uses a multi-task learning process to address the detection issue by combining classification and bounding box regression. The system normally includes two stages: the RPN and Faster R-CNN header network, which uses a convolutional backbone to extract high-level features from pictures. The Faster R-CNN method replaces the selective search scheme (Ren *et al.*, 2015), originally used in RPN. For the Faster R-CNN, in its first stage, the RPN shifts a convolutional sliding window over the feature maps to produce proposals. Multi-scale anchors are used at each point on the feature map to forecast multi-candidate object boxes, and the top-ranked object candidates are cropped using an RoI pooling layer derived from the feature extractor's intermediate layer. The latter deals with the varied size issue in the feature maps. Finally, the candidate boxes are sent to the fully connected layer. In the second stage, each proposal undergoes a final classification and box-refinement procedure (Sultana *et al.*, 2020; Du *et al.*, 2020).

The experimental result reveals that Faster R-CNN and SSD can deliver better detection accuracy than YOLO, although the latter is faster than SSD and Faster R-CNN. Considering the recognition rate and accuracy, we use the positive-match SSD prediction model to implement pallet recognition using RGB images in the first stage of PILA (Du, 2020; Zaccaria *et al.*, 2020). As shown in equation (1), if the IoU score exceeds 0.5, the matching value m is 1, i.e. a positive match. On the contrary, a zero or negative match means that the object is detected or not. The SSD model's initial learning rate and batch size are set to 0.004 and 24:

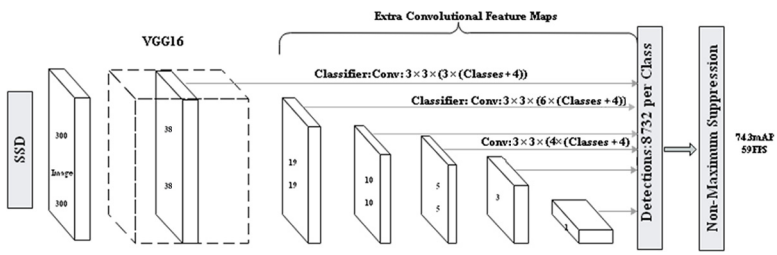
$$m = \begin{cases} 1 \rightarrow Positive\ matches & if \quad IoU > 0.5 \\ 0 \rightarrow Negative\ matches & if \quad IoU \leq 0.5 \end{cases} \tag{1}$$

Because pallets are often placed randomly in the warehouse environment, the robot must handle the multiple pallets presented in its field of view. Thus, images with a complex background are collected during the data set creation to enhance the model's robustness and compliance. All images are converted to 300 × 300 pixels, which are then input into the SSD model. The pallets within the image are manually labelled as ground truth boxes before training.

3.2 Pallet RGB-Depth data set

As there are many different standards in the industry, such as ISO, European and North American, pallets come with various materials, sizes and shapes, and thus, a representative training

Figure 4 The diagram of SSD architecture



J_ID: IR ART NO: 10.1108/IR-05-2022-0123 Date: 28-December-22 Page: 5 Total Pages: 13 4/Color Figure(s) ARTTYPE="ResearchArticle"

data set is necessary for general usage. The plastic and wood pallets are the most common and intensely tested during the verification. Figure 5 shows, more than five types of pallets are collected in our data set under different scenarios and conditions, including cases where the pallet is on the rack, on the ground, with a card box or with small viewpoint angles. Furthermore, the pallet images of different lighting conditions, floor conditions and partial occlusion are included to match the real environment in a warehouse. Additionally, there are two- and four-way pocket pallets, which allow the forklift to pick from two-way or four-way directions. Therefore, we have collected both types in the dataset to enhance the model's generalization ability. The pallet assembly information is listed in Table 3.

More than 1,000 pallet pictures are collected and used to train the model. As the valid RGB and depth images are taken at the forklift's operation distance, the DNN results involving a potential pallet area are further filtered with a sanity check based on the minimum and maximum area limits and length-to-width ratio. The recognition rates of the three pallet types are illustrated in Table 4. The average detection rate exceeds 98%, demonstrating great potential for warehouse use. Examples of the pallet detection results with a labelled pallet RoI are illustrated in Figure 6. Regardless of the presence of a card box, several pallets in the image or tilted pallets, these can be easily distinguished. The overall results are discussed in Section 3.

Figure 5 Various types of pallets used in the training dataset

Notes: Pallets on the ground, pallets with the cardbox, on the racks, and tilting pallets are among the multiple types of pallets

Table 3 Information about some pallet types in the data set

Pallet types	Material	Colour	Dimensions (W × L × H) mm
Two-way entry	Wood	Wooden	700 × 1,400 × 130
Four-way entry	Wood	Wooden	800 × 1,200 × 145
	Plastic	Blue	1,000 × 1,200 × 150
	Plastic	Blue	1,219 × 1,143 × 140
	Plastic	White	914 × 1,200 × 150

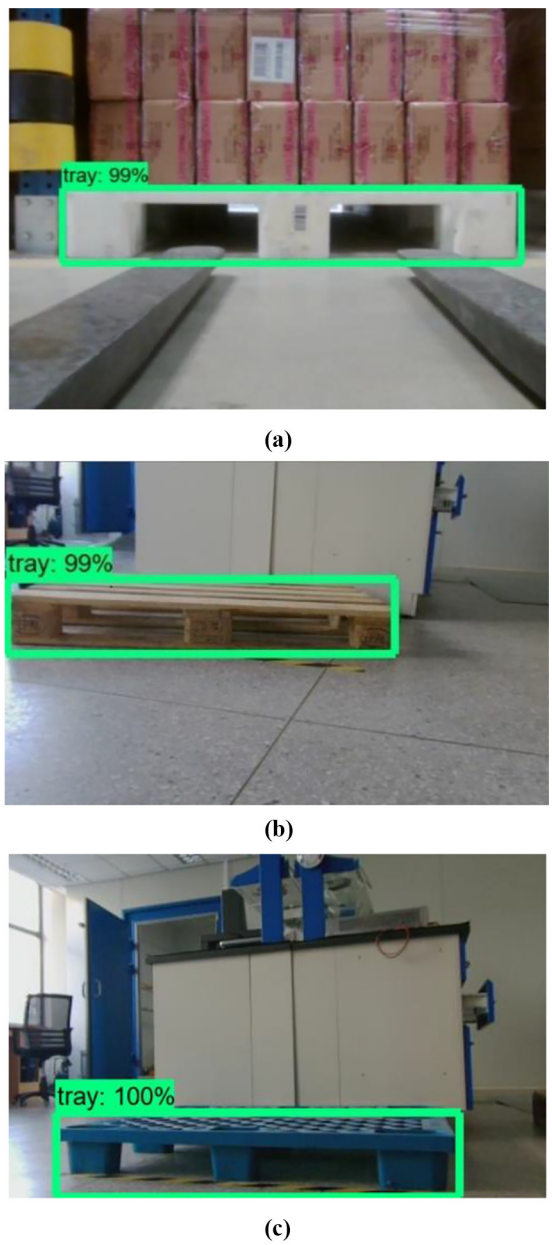
Table 4 Pallet recognition results of SSD model

Pallet material	Colour	Dimensions (W × L × H) mm	Recognition rate (%)
Wood	Wooden	700 × 1,400 × 130	98
Plastic	White	914 × 1,200 × 150	99
Plastic	Blue	1,000 × 1,200 × 150	98.50

3.3 RGB and point cloud data fusion

After obtaining the output from the SSD model, the detection bounding box of the RGB image can be further selected by removing the small-area pallets, highly-tilted pallets or duplicated pallets. Then, the RGB and depth images from the RGB-D camera must be aligned in space using the camera projection principle. Hence, we calibrate the camera using a chessboard (Zhang, 2000) to obtain the intrinsic and extrinsic parameters reported in Tables 5, 6 and 7, where f_x, f_y, c_x and c_y are the focal length and the principal point represented in a 3×3 matrix [equation (2)], respectively. The distortion

Figure 6 RoI of pallet images



Notes: (a) The scene of a pallet in the field of vision during detection; (b) The tilted wooden pallet; (c) The tilted plastic pallet

J_ID: IR ART NO: 10.1108/IR-05-2022-0123 Date: 28-December-22 Page: 6 Total Pages: 13 4/Color Figure(s) ARTTYPE="ResearchArticle"

parameters for each camera are listed as $[k_1, k_2, k_3, k_4, k_5, k_6]$ for radial distortion and $[p_1, p_2,]$ for tangential distortion, as shown in equations (3) and (4) (Villena-Martínez *et al.*, 2017). The extrinsic parameters, 3×3 rotation matrix $r = [(r_{11}, r_{21}, r_{31})^T, (r_{12}, r_{22}, r_{32})^T, (r_{13}, r_{23}, r_{33})^T]$, and a 3×1 translation vector $t = (t_x, t_y, t_z)^T$ in Table 8 are used to convert the world coordinate system to the camera coordinate system:

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$\begin{aligned} i_c &= i(1 + k_1 r + k_1 r^2 + k_3 r^3 + k_4 r^4 + k_5 r^5 + k_6 r^6) \\ j_c &= j(1 + k_1 r + k_1 r^2 + k_3 r^3 + k_4 r^4 + k_5 r^5 + k_6 r^6) \end{aligned} \tag{3}$$

$$\begin{aligned} i_c &= i + \left[2p_1 y + p_2(r^2 + 2x^2)\right] \\ j_c &= j + \left[p_1(r^2 + 2y^2) + 2p_2 x\right] \end{aligned} \tag{4}$$

where (i, j) is the uncalibrated position of a pixel in the image, r is the radius to the principal point and (i_c, j_c) is the calibrated position.

3.4 Point cloud processing

In this section, the RGB and depth data fusion is further processed, as outlined in Algorithm 1 (Rusu *et al.*, 2011). The processing mainly includes filtering out data points that do not belong to the pallet surface and calculating key points based on selected geometric rules.

Algorithm 1: Point cloud processing

1. Convert the depth image to Point Cloud data.
2. Remove out-of-range and scattered Point Cloud outliers.
3. Down-sample the Point Cloud.
4. Segment front surface planes from Point Cloud data.
5. Project filtered inliers of Point Cloud data to form a 2D plane.
6. Extract the x and y line group from selective rules.
7. Pick the best x and y-line candidates to form a "T-shape."
8. Determine triangle-centric points of the pallet's front-face.

3.4.1 Point cloud filtering

The point cloud filtering process comprises the three steps illustrated in Figure 7. Firstly, the point cloud data are input into a pass-through filter to preserve all points with a z value (distance) ranging between 0.5 m and 3 m. Then outliers are removed, and the surface data is averaged. Finally, down-sampling is applied to reduce the computational cost.

3.4.2 Vertical plane extraction:

It is feasible to assume that the pallet surface is vertically facing the camera. The pipeline extracts the pallet's vertical plane illustrated in Figure 7, where several 2D plane candidates are extracted from the projecting filtered inliers along the z direction after performing point cloud segmentation. A possible plane can be found based on the centroid score, and the object's centroid is given by equation (5), where m_i is set to 1. A simplified version of that formula is given in equation (6). The centroids from multiple planes are compared, and the closest to the view centre is the most likely vertical plane of interest. Finally, the angle of the pallet's front face to the camera is calculated:

$$P_c = \frac{1}{M} \sum_{0}^{n} m_i r_i \tag{5}$$

where $r_i = (x_i, y_i, z_i)$, $i = 1,2,\ldots,n$, with n denoting the coordinate of each point, and m_i represents the mass of the corresponding point, which is set to 1 (Figure 9):

$$P_c = \frac{1}{n}\left(\sum_{i=0}^{n} x_i, \sum_{i=0}^{n} y_i, \sum_{i=0}^{n} z_i\right) \tag{5}$$

Table 7 Calibrated RGB and depth camera parameters of the tangential distortion coefficient (TDC)

TDC	p_1	p_2
RGB camera	0.00088	−0.00055
Depth camera	−0.00096	−0.00051

Table 8 Calibrated RGB-D camera parameters of the rotation and translation

$(r_{11}, r_{21}, r_{31})^T$	$(r_{12}, r_{22}, r_{32})^T$	$(r_{13}, r_{23}, r_{33})^T$	$(t_x, t_y, t_z)^T$
0.9999	−0.0027	0.0043	47.8448
0.0027	0.9999	0.0007	−0.0501
−0.0043	−0.0007	0.9999	−2.2684

Table 5 Calibrated RGB and depth camera parameters of the focal length and principal point (pixel)

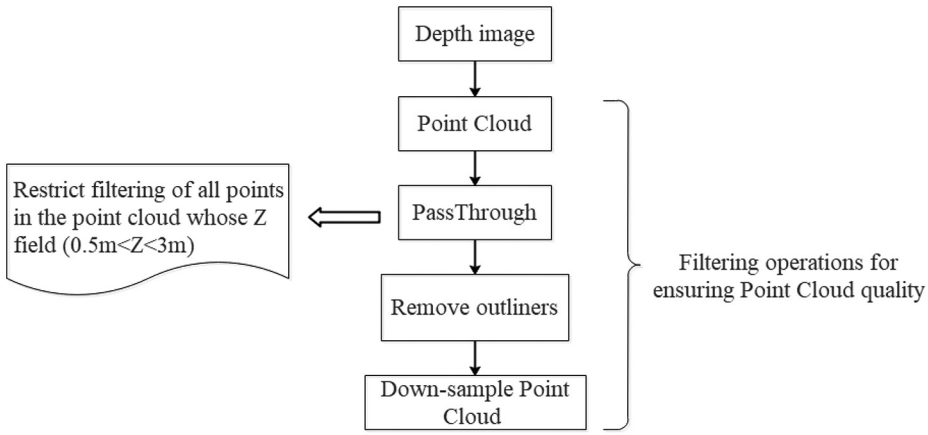
	Focal length x (f_x)	Focal length y (f_y)	Principal point x (c_x)	Principal point y (c_y)
RGB camera	528.328	528.077	311.297	187.764
Depth camera	626.491	626.088	316.583	235.191

Table 6 Calibrated RGB and depth camera parameters of the radial distortion coefficient (RDC)

RDC	1st-order (k_1)	2st-order (k_2)	3st-order (k_3)	4st-order (k_4)	5st-order (k_5)	6st-order (k_6)
RGB camera	−0.160	0.189	−0.053	0.246	0.062	0.012
Depth camera	−0.124	0.117	−0.039	0.132	0.034	0.013

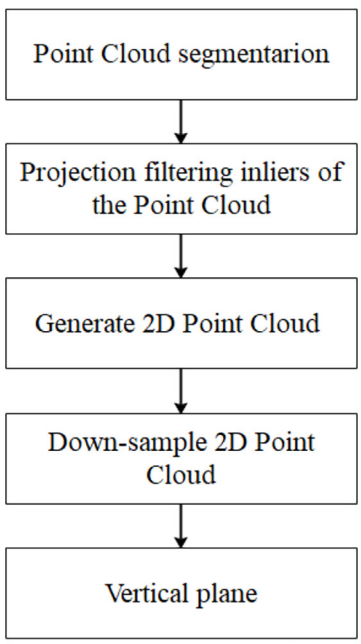
J_ID: IR ART NO: 10.1108/IR-05-2022-0123 Date: 28-December-22 Page: 7 Total Pages: 13 4/Color Figure(s) ARTTYPE="ResearchArticle"

Figure 7 Pipeline of point cloud filtering



Notes: Firstly, the point cloud data converted from depth image is performed through a pass-through filter to restrict filtering of all points in the point cloud whose Z field is greater than 0.5 m and less than 3 m. Then outliners are removed, and point cloud is downsampled

Figure 8 Pipeline of extracting vertical plane



Notes: After point cloud segmentation, the filtering inliers of the point cloud are projected, and the 2D point cloud is generated. Finally, the 2D point cloud is downsampled, and then a vertical plane is generated

3.4.3 Lines extraction and pallet localization

In the last part of PILA, the x- and y-lines (pallet's horizontal and vertical edges, respectively) of the "T-shape" and the pallet centre points A, B and C are calculated. Firstly, the boundary points are detected by slicing the 2D plane horizontally and vertically. During the horizontal slicing, the vertical part of the boundary contour is formed by the points that are either far from the left or the right neighbour points. A similar horizontal part of the boundary contour is retrieved by vertical slicing. Then, the "T-shape" is found based on the combination of the bottom line of the pallet top (x-lines) and the outside boundary of the middle post (y-lines), as illustrated in Figure 3. The first step in detecting the "T-shape" is to determine the x-line and the two y-lines close to the geometric centre of the RoI. The pipeline of the line extraction and pallets location decision process is presented in Figure 9. The horizontal boundary and vertical boundary points in the x and y directions are extracted, and the selected x-line length must be longer than 1 m, and y-lines must be longer than 10 cm to reduce false results. The valid x-line and y-line should have a good amount of points representing valid pallet edges, i.e. 150 points and 8 points for the x-line and y-line, respectively, Figure 10. These parameters are determined experimentally from the RGB-D camera resolution at the working distance. After sorting all x and y lines, one x-line and two y-lines closest to the geometric centre point form the two intersection points, A and B. The distance between the pallet and the camera is then calculated using point C as the geometric centre point of the tray plane.

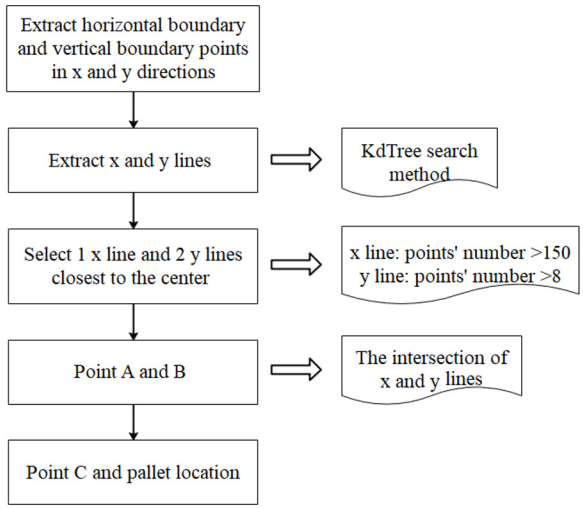
Figure 11 depicts the four basic PILA steps graphically, where (a) and (b) are the RGB images and point cloud data, and (c) and (d) are generated through the pallet identification and point cloud processing to locate the pallet centre.

4. Vehicle alignment and approach algorithm

The forklift robots must align the pallet before conducting the picking operation. Hence, we propose a two-phase vehicle alignment process to gradually approach the pallet location and adjust the angle to achieve a precise vehicle alignment. In the first phase, the robot adjusts the angle and the distance to allow the forklift to align the arms to the pallet's centre. A straight path is followed in the second phase, and the angle is adjusted slightly within the "maxorient" to approach the pallet.

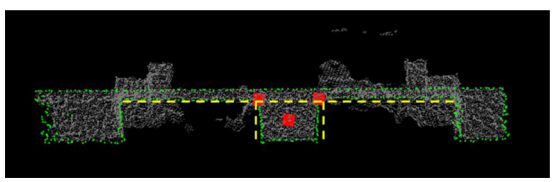
J_ID: IR ART NO: 10.1108/IR-05-2022-0123 Date: 28-December-22 Page: 8 Total Pages: 13 4/Color Figure(s) ARTTYPE="ResearchArticle"

Figure 9 The pipeline of lines extraction and pallets location



Notes: The horizontal boundary and vertical boundary points in x and y directions are extracted first. x and y lines by KdTree search method are executed, and 1 of x-line and 2 of y lines are picked closest to the centre

Figure 10 T-shape of pallet surface in the view of point cloud data. The *x*-line and *y*-line pairs in yellow are the lines selected by the algorithm



F12 In Figure 12, d is the distance of the segment that connects the robot and the pallet centres, and h and b are the vertical and horizontal distances from the robot to the pallet centres, respectively. The parameter θ ($\theta = \varepsilon$) represents the angular relationship between h and b, and ϕ is the robot's orientation to the perpendicular line presented in yellow. Parameters d, b and ε are obtained from PILA. The suggested algorithm is summarized in Algorithm 2.

Algorithm 2: Vehicle alignment approach algorithm
Initializing: Obtain the pose of the robot concerning the centre of the pallet and convert data to the orientation of ϕ.
First phase: Control the wheels to reach ϕ (the robot's orientation). Correct the orientation until the angle becomes lower than θ_{min}.
Pre-second phase: Rotate the robot as long as it points to the centre of the pallet and the orientation is close to zero.
Second phase: Follow the straight path to the pallet, make the distance and orientation not exceed the limits until the robot reaches the pallet's centre and the picking operation is completed.

5. Results and analysis

5.1 Experimental setup

The experiment aims to verify PILA's performance on pallet localization under various conditions. Thus, a Pico Zense RGB-D camera is placed between the two forklift arms to capture RGB and depth images. The NUC unit runs a 64-bit windows embedded system equipped with an external CAN device to transmit/receive messages. The experimental setup is illustrated in Figure 13(b). In the following experiments, we analyse the F13 impact of the pallet angle, the distance from the forklift to the pallet, and the various vertical and horizontal offsets.

5.2 Pallet detection and localization results

The ground truth table and the experimental results of the pallet location are reported in Table 9. The horizontal displacement, the T9 vertical displacement, the z distance and the angle between the camera and the pallet are represented as x, y, z and angle values in Table 9. Section 1 in the table presents the measurements at six different distances from 1.25 m to 3 m without any offset. Section 2 reports that the horizontal and vertical offsets are added at a distance of 1.5 m, and Section 3 presents the measurement with the angle offsets at a distance of 1.25 m and 1.5 m.

The absolute error, mean absolute error (MAE) and standard deviation (STD) of the absolute error are reported in Table 10. T10 The results highlight that the maximum absolute error of the horizontal and vertical displacement, distance and angle is 9.19 mm, 9.77 mm, 11.65 mm and 0.65 degrees, respectively. The minimum absolute errors are 0.74 mm, 0.08 mm, 0.59 and 0.07 degrees, respectively. Additionally, the STD values indicate that the measurement errors of the x, y, z and angle are consistent in the operating range. Figure 14 reveals that the absolute error F14 values of x, y and z in Section 1 of Table 10 increase with a distance from 1.25 m to 3 m. However, the absolute error curve of the angles is relatively stable with distance, i.e. PILA's angular estimation is more resistant to errors than the x, y and z location. To the best of our knowledge, this is one of the most accurate results of pallet localization for industrial robot applications.

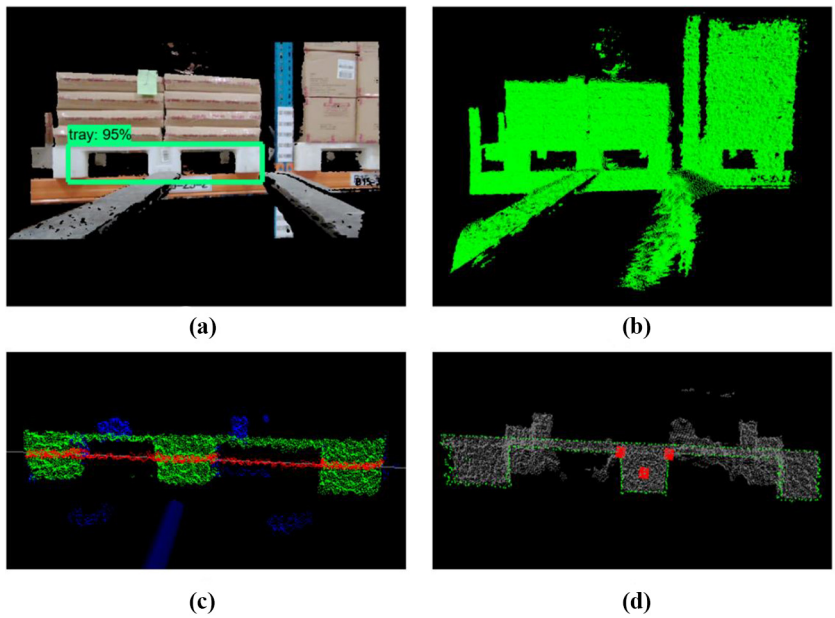
The experimental results show that for a range of up to 3 m, the proposed PILA algorithm can identify and locate pallets in 3D and the average absolute errors are in the order of 1 cm or within 0.65 degrees. The average time for each test is around 700 ms, meeting the "real-time" requirements. The experimental results highlight that PILA can use RGB imagery to obtain a stable and high recognition accuracy based on two functional stages, and the depth data guarantee a very high localization accuracy. By integrating a vision-based and range-based method, we exploit the advantages of RGB and depth information, enhancing our approach's robustness and affording its warehouse deployment.

5.3 Vehicle alignment algorithm results

Due to the extreme effort required to implement the VAA algorithm in a real forklift robot, we investigate VAA's performance in the gazebo simulation instead, with the corresponding results presented in Table 11. In the simulations, PILA's output is assigned to the VAA as the alignment target, and a differential-driven forklift model is used to simulate the forklift's behaviour. The tests consider three cases: different distances, initial angles and horizontal offset. The robot is set to various initial positions with different angles, and the PILA node sends the target positions to the VAA node. After completing the

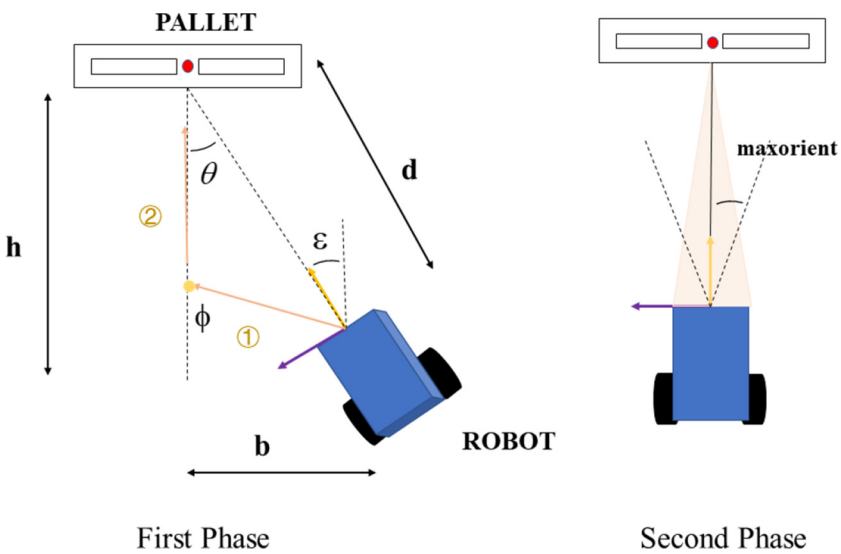
J_ID: IR ART NO: 10.1108/IR-05-2022-0123 Date: 28-December-22 Page: 9 Total Pages: 13 4/Color Figure(s) ARTTYPE="ResearchArticle"

Figure 11 Graphic presentation of four primary steps of PILA



Notes: (a) The RGB image of pallet; (b) the raw point cloud data converted from depth image; (c) the filtered point cloud data according to pallet recognition; (d) final point cloud data for pallet location

Figure 12 The VAA diagram of the forklift robot being aligned with the pallet



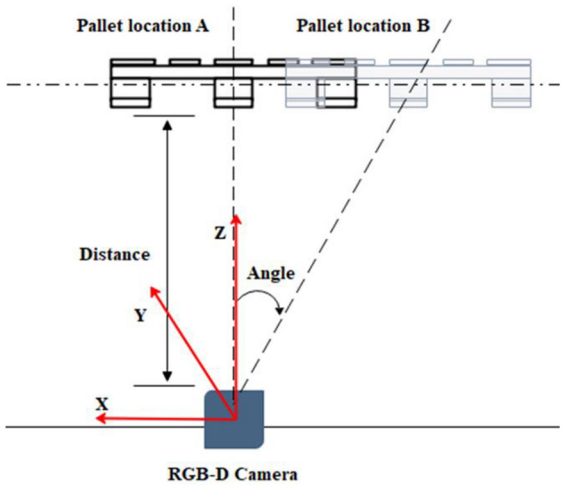
alignment operation, the final distance and angle to pallet results are reported in Table 11, with desired distance and angle to the target set as 100 mm and 0 degrees, respectively. The results reveal that the robot can adjust the angle and approach a pallet as desired. The average final distance is 101.24 mm with 2.34 mm STD, and the average facing angles are 0.81 degrees with 0.39 degrees STD. This proves that PILA and VAA effectively collaborate to assist in identifying and picking up the pallet with additional displacements or angle offsets.

6. Conclusion and discussion

This paper presents the logistic PILA and VAA algorithms. PILA combines vision-based recognition and point cloud processing to recognize and accurately localize pallets within a reasonable runtime. Specifically, RGB images are treated by a DNN, and RGB and depth image fusion techniques are used to align the RGB image with the depth image in the pallet area. Then, the "T-shape" in the pallet centre is located to afford

J_ID: IR ART NO: 10.1108/IR-05-2022-0123 Date: 28-December-22 Page: 10 Total Pages: 13 4/Color Figure(s) ARTTYPE="ResearchArticle"

Figure 13 A diagram of experimental setup (a) and forklift system used in the pallet picking experiment



(a)

(b)

Table 9 The comparison of practical and computed localization results (mm and degree)

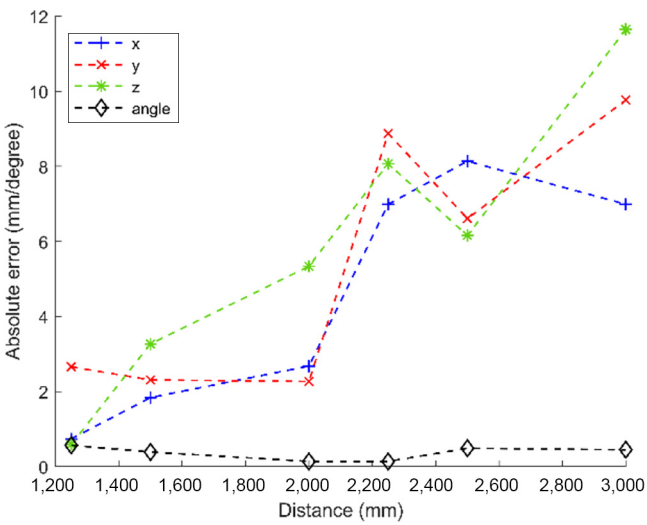
	Ground truth positions				Estimated results			
	x	*y*	*z*	Angle	*x*	*y*	*z*	Angle
Section 1	0	0	1,250	0	0.74	2.66	1,249.41	0.57
	0	0	1,500	0	1.84	2.31	1,503.27	0.39
	0	0	2,000	0	2.68	−2.27	2,005.34	−0.14
	0	0	2,250	0	4.76	−8.88	2,241.93	0.14
	0	0	2,500	0	8.14	6.61	2,493.84	−0.49
	0	0	3,000	0	6.99	−9.77	3,011.65	−0.45
Section 2	10	0	1,500	0	13.87	6.82	1,496.78	0.54
	50	0	1,500	0	53.19	2.39	1,496.68	0.65
	100	0	1,500	0	107.26	1.65	1,492.43	0.51
	0	5	1,500	0	−6.27	6.36	1,489.55	0.48
	0	10	1,500	0	9.19	10.08	1,497.52	0.45
	0	15	1,500	0	−4.06	17.13	1,492.85	0.13
Section 3	0	0	1,250	5	2.66	1.47	1,242.53	5.33
	0	0	1,250	10	6.11	−1.51	1,252.45	9.68
	0	0	1,250	15	4.04	−3.78	1,256.14	14.59
	0	0	1,500	5	3.35	−5.37	1,498.44	4.93
	0	0	1,500	10	−1.79	−3.84	1,494.38	9.67
	0	0	1,500	15	3.49	−0.21	1,509.67	14.87
	10	0	1,500	5	12.33	−1.9	1,500.99	4.66
	50	0	1,500	10	50.9	0.99	1,503.86	10.28
	100	0	1,500	15	103.23	0.29	1,499.21	16.01

AQ: 14

Table 10 Absolute error, MAE (mm and degree) and STD of absolute error

	Absolute error			
	x	*y*	*z*	Angle
Section 1	0.74	2.66	0.59	0.57
	1.84	2.31	3.27	0.39
	2.68	2.27	5.34	0.14
	6.99	8.88	8.07	0.14
	8.14	6.61	6.16	0.49
	6.99	9.77	11.65	0.45
Section 2	3.87	6.82	3.22	0.54
	3.19	2.39	3.32	0.65
	7.26	1.65	7.57	0.51
	6.27	1.36	10.45	0.48
	9.19	0.08	2.48	0.45
	4.06	2.13	7.15	0.13
Section 3	2.66	1.47	7.47	0.33
	6.11	1.51	2.45	0.32
	4.04	3.78	6.14	0.41
	3.35	5.37	1.56	0.07
	1.79	3.84	5.62	0.33
	3.49	0.21	9.67	0.13
	2.33	1.9	0.99	0.34
	0.9	0.99	3.86	0.28
	3.23	0.29	0.79	1.01
MAE	4.24	3.15	5.13	0.39
STD	2.44	2.85	3.17	0.18

Figure 14 Absolute error curve of *x*, *y*, *z* and angle with the distance changing from 1.25 m to 3 m



PILA's application on multiple pallet types. The experimental results indicate that PILA affords a 90% pallet recognition rate, and the MAE of pallet localization and angle is less than 1 cm or 0.4 degrees. Additionally, VAA is proposed to implement robot motion planning using the PILA output to fulfil the pallet-picking operation. All algorithms have been implemented in C++ for practical pallet-picking scenarios, demonstrating excellent efficiency considering precision, speed and working distance. Compared with most solutions based on computer vision or depth image, PILA has been proven feasible for forklift robot applications in the logistics warehouse.

Based on our observation, the "T-shape" geometrical rule can work as a loose regulation to locate the pallet's centre without knowing its size. This strategy is more meaningful for the massive deployment of intelligent logistics, affording forklifts and AGV robots to detect and locate pallets with random offsets.

6.1 Future work

It should be mentioned that the PILA performance will drop quickly beyond the distance of 3 m because the depth image quality of the RGB-D camera drops with distance. This can be improved in future works by choosing better depth-measuring hardware. In the future, more complex environmental cases, such as occlusion by humans or goods, defective pallets and various lighting conditions, must be considered. Moreover, the PILA algorithm must be fully implemented with VAA in a "real" forklift robot and verified in a warehouse environment. AQ: 12

References

Behrje, U., Himstedt, M. and Maehle, E. (2018), "An autonomous forklift with 3d time-of-flight camera-based localization and navigation", *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. pp. 1739-1746, *IEEE*.

Casado, F., *et al.* (2017), "Pose estimation and object tracking using 2d images", *Procedia Manufacturing*, Vol. 11, pp. 63-71.

Deshpande, A.M., Telikicherla, A.K., Jakkali, V., *et al.* (2020), "Computer vision toolkit for non-invasive monitoring of factory floor artifacts", *Procedia Manufacturing*, Vol. 48, pp. 1020-1028.

Du, L., *et al.* (2020), "Overview of two-stage object detection algorithms", *Journal of Physics: Conference Series*, Vol. 1544 No. 1.

Fontana, E., Rizzini, D.L. and Caselli, S. (2021), "A combinatorial approach to detection of box pallet layouts", *2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP)*. pp. 417-422, *IEEE*.

Iinuma, R., Kojima, Y., Onoyama, H., *et al.* (2020), "Pallet handling system with an autonomous forklift for outdoor fields", *Journal of Robotics and Mechatronics*, Vol. 32 No. 5, pp. 1071-1079.

Jia, F., Tao, Z. and Wang, F. (2021a), "Wooden pallet image segmentation based on Otsu and marker watershed", *Journal of Physics: Conference Series*, Vol. 1976 No. 1, p. 012005, IOP Publishing.

Jia, F., Tao, Z. and Wang, F. (2021b), "Pallet detection based on Halcon for warehouse robots", *2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID)*, pp. 401-404, doi: 10.1109/AIID51893.2021.9456540.

Jiao, L., *et al.* (2019), "A survey of deep learning-based object detection", *IEEE Access*, Vol. 7, pp. 128837-128868.

Li, T., Huang, B., Li, C., *et al.* (2019), "Application of convolution neural network object detection algorithm in logistics warehouse", *the Journal of Engineering*, Vol. 2019 No. 23, pp. 9053-9058.

Mohamed, I.S., *et al.* (2020), "Detection, localisation and tracking of pallets using machine learning techniques and 2D range data", *Neural Computing and Applications*, Vol. 32 No. 13, pp. 8811-8828.

Mok, C., Baek, I., Cho, Y.S., *et al.* (2021), "Pallet recognition with Multi-Task learning for automated guided vehicles", *Applied Sciences*, Vol. 11 No. 24, p. 11808.

Molter, B. and Fottner, J. (2018), "Real-time pallet localization with 3d camera technology for forklifts in logistic environments", *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, pp. 297-302, *IEEE*.

Redmon, J., *et al.* (2016), "You only look once: unified, real-time object detection", *Proceedings of the IEE Conference on Computer Vision and Pattern Recognition*. pp. 779-788.

Ren, S., *et al.* (2015), "Faster R-CNN: towards real-time object detection with region proposal networks", *Advances in Neural Information Processing Systems*, Vol. 28, pp. 91-99.

Rusu, R., Bogdan. and S., Cousins. (2011), "3d is here: point cloud library (PCL)", *2011 IEEE international conference on robotics and automation*. pp. 1-4, *IEEE*.

Song, Q., *et al.* (2020), "Dynamic path planning for unmanned vehicles based on fuzzy logic and improved ant colony optimization", *IEEE Access*, Vol. 8, pp. 62107-62115.

Srinidhi, C.L., Ciga, O. and Martel, A.L. (2021), "Deep neural network models for computational histopathology: a survey", *Medical Image Analysis*, Vol. 67, p. 101813.

Sultana, F., Sufian, A. and Dutta, P. (2020), "A review of object detection models based on convolutional neural network", *Intelligent Computing: Image Processing Based Applications*, pp. 1-16.

Syrjänen, A. (2021), "Experimental evaluation of depth cameras for pallet detection and pose estimation[D]".

Tsiogas, E., Kleitsiotis, I., Kostavelis, I., *et al.*, 2021 "Pallet detection and docking strategy for autonomous pallet truck AGV operation", *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, *IEEE*, pp. 3444-3451.

Villena-Martínez, V., *et al.* (2017), "A quantitative comparison of calibration methods for RGB-D sensors using different technologies", *Sensors*, Vol. 17 No. 2, p. 243.

Walter, M.R., Karaman, S., Frazzoli, E., *et al.* (2010), "Closed-loop pallet manipulation in unstructured environments", *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 5119-5126, *IEEE*.

Wang, S., *et al.* (2021), "A lightweight localization strategy for LiDAR-Guided autonomous robots with artificial landmarks", *Sensors*, Vol. 21 No. 13, p. 4479.

Ward, R., Soulatiantork, P., Finneran, S., *et al.* (2021), "Real-time vision-based multiple object tracking of a production process: industrial digital twin case study", *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, Vol. 235 No. 11, pp. 1861-1872.

Wenhan, W.U., Ming, Y., Bing, W., *et al.* (2019), "Pallet detection based on contour matching for warehouse robots", *Journal of Shanghai Jiaotong University*, Vol. 53 No. 2, p. 197.

Xiao, J., *et al.* (2017), "Pallet recognition and localization using an RGB-D camera", *International Journal of Advanced Robotic Systems*, Vol. 14 No. 6, p. 1729881417737799.

Zaccaria, M., *et al.* (2020), "A comparison of deep learning models for pallet detection in industrial warehouses", *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*. pp. 417-422, *IEEE*.

Zhang, Z. (2000), "A flexible new technique for camera calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22 No. 11, pp. 1330-1334.

Zou, Z., *et al.* (2019), "Object detection in 20 years: a survey", arXiv preprint arXiv:1905.05055.

Further reading

Liu, W., *et al.* (2016), "SSD: single shot multibox detector", *European Conference on Computer Vision*, *Springer*, *Cham*, pp. 21-37.

About the authors

Yongyao Li, master student in 2019's class and in the joint program of Changchun University of Science and Technology and Suzhou Institute of Biomedical Engineering and Technology (SIBET). His research topics are focused on image pattern detection based on neural networks and 3D point cloud processing for Biomedical applications.

Guanyu Ding, received his master degree in Electrical Engineering from the University of Texas, Austin, in 2012. His research interest includes electrical circuits, automatic control theory, machine vision and machine learning, with applications to autonomous vehicles and robot.

Chao Li, received a master in Electrical Engineering and Computer Science from the University of California, Irvine. Now she is serving as the technical director at Pilot AI company. Her work involves the hardware development of Robot system and self-driving vehicles.

Sen Wang, a master student in the 2019's class and in the joint program of Changchun University of Science and Technology and Suzhou Institute of Biomedical Engineering and Technology (SIBET). His research topics are focused on navigation and path planning.

Qinglei Zhao obtained a bachelor's degree in mechanical engineering and automation from Tianjin University in 2005, a master's degree in software engineering from the Chinese University of Science and Technology in 2010, and a doctorate in mechanical and electronic engineering from the University of Chinese Academy of Sciences in 2016. He has been engaged in the research of embedded system design and automatic control at Changchun Institute of Optics, Fine Mechanics and Physics of the Chinese Academy of Sciences.

Qi Song received his PhD in Electrical Engineering and Computer Science from the University of California, Irvine, in 2013 and is now the Professor at Suzhou Institute of Biomedical Engineering and Technology, Chinese academy of sciences. His research interest includes a SLAM navigation system, the intersection of control theory, machine vision and machine learning, with applications to autonomous vehicles and robot. Qi Song is the corresponding author and can be contacted at: songq@sibet.ac.cn

AQ: 13