Robotics and Computer–Integrated Manufacturing 73 (2022) 102229

Contents lists available at ScienceDirect

Robotics and Computer-Integrated Manufacturing

journal homepage: www.elsevier.com/locate/rcm

Full length article

Learning-based object detection and localization for a mobile robot manipulator in SME production

Zhengxue Zhou, Leihui Li, Alexander Fürsterling, Hjalte Joshua Durocher, Jesper Mouridsen, Xuping Zhang *

Aarhus University, Department of Mechanical and Production Engineering, Inge Lehmanns Gade 10, Aarhus, 8000, Aarhus, Denmark

ARTICLE INFO

Keywords:
Object detection
Mobile manipulator
Localization
SME production

ABSTRACT

Increasing research attention has been attracted to automatic production processes in small and medium-sized enterprises (SMEs) using collaborative robotic systems. In this work, we develop an object detection solution based on deep learning on 3D point clouds for a collaborative mobile robot manipulator to automate SME production. In this solution, a 3D point cloud technology is adopted to measure the shape and depth information of targeted objects in SME production, for instance, name tags production and plug-in charging. Deep learning is then employed to deal with the uncertainty in 3D detection, such as inconsistent light conditions and the irregular distribution and structural ambiguity of point clouds. A 2D camera is employed to calibrate the relative positions of the mobile manipulator to workstations. The mobile robot manipulator is equipped with cameras, an in-house developed adaptive gripper, and a learning-based computer vision system developed in this work. The principle and procedures of the proposed 3D object detection and 2D calibration are presented in detail. The automatic name tags production and plug-in charging experiments are conducted to validate the object detection, localization algorithms, and tools developed and employed in production cases using the mobile robot manipulator.

1. Introduction

During the past decades, industrial robots have been increasingly integrated into manufacturing across industries and are in some cases vital for the companies' market competitiveness [1,2]. However, this development is generally limited to larger companies due to the costs, spaces, related techniques, and the fact that small and medium-sized enterprises (SMEs) usually run smaller product batch sizes to meet the demands of a diversified customer base [3]. They cannot afford large and complex robots because of limited space, lack of technical staffs, and changeable working environments such as layout and light condition. Such dynamic environments can be particularly challenging for applying robots [4]. Therefore, the collaborative robots, which are lightweight and space-saving, have multiple uses, and are safe to collaborate with human workers, are necessary for automating the production in SMEs. On the other hand, a mobile manipulator can move freely around on a production site to perform various tasks at different workstations in changeable working environments, make robotic implementations more agile, and create new cost-beneficial opportunities [5]. There are still lots of challenges for the application of such mobile manipulators in SMEs. In this work, we are devoted to developing the solutions to the three challenges: (1) Detecting workstations and humans with the increased robustness in the SME dynamic working environments; (2) Localizing and navigating the mobile manipulator with respect to workstations precisely; (3) Developing adaptive tools and grippers to undertake stable and precise manipulation for diverse production tasks.

Various research efforts have been undertaken to explore the application of mobile robot manipulators to achieve production tasks in factories. However, most of the efforts were on simulation work or experimental testing conducted in laboratory environments to simulate production tasks using mobile manipulators. For example, Zhou, etc. [6] developed a mobile manipulator, which can collaborate with humans in the aerospace manufacturing industry. However, the system was not validated by experiments. Several other mobile manipulators were employed and tested in laboratories, for example, the construction industry [7], automatic order-picking [8], aircraft canopy polishing [9], and automotive wiring harness assembly [10]. Nevertheless, the above studies only conducted the testing of mobile manipulators in laboratory environments. Only a few mobile manipulator applications were tested in real industrial environments. For example, a mobile manipulator was applied to pick and place objects between different workstations in a

* Corresponding author.
E-mail address: xuzh@mpe.au.dk (X. Zhang).

https://doi.org/10.1016/j.rcim.2021.102229
Received 8 February 2021; Received in revised form 7 May 2021; Accepted 7 July 2021
Available online 17 August 2021

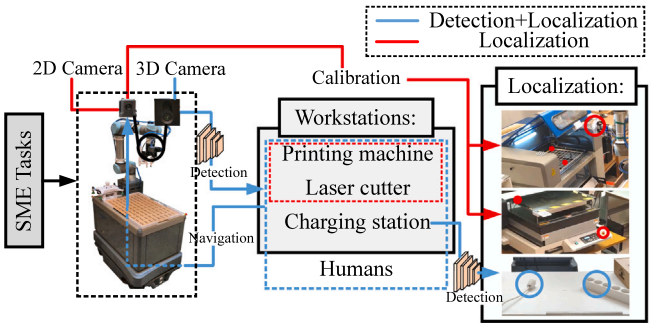

Fig. 1. Robotic solution for automatic production and self-charging tasks.

factory [11]. Ultrasonic range finders were used for detecting obstacles by finding the range and bearing of the obstacles. Another mobile manipulator, rob@work [12], was applied to assist humans with the assembly of hydraulic pumps and the gas metal arc welding. Hvilshøj, Madsen, and Andersen, etc. [13–17] also evaluated autonomous industrial mobile manipulators in pump manufacturer Grundfos' factory to solve the task of producing rotors.

It is well known that a collaborative mobile manipulator is considered as an effective solution to the automation of the high-mix, low-volume production by navigating among different workstations. Most mobile robots navigate based on pre-established maps built by simultaneous localization and mapping (SLAM), which has the problem of the robustness to dynamic environments when the layout nearby targeted objects is changed [18]. Consequently, one of the critical challenges is to develop a robust detection algorithm for the mobile manipulator to navigate among different workstations with the ignorable impact of the changed layout. Calculating the precise relative position of targeted objects to the mobile manipulator is another challenging issue. On the other hand, SMEs usually have customized products in small batch sizes. Therefore, it is imperative to develop an adaptive robotic gripper and tools capable of handling diverse production tasks.

Object detection is a routine operation in industrial production using robots and an issue of great importance in autonomous driving. A variety of methods can be utilized for the object detection of mobile manipulators and robots, such as electromagnetic sensors [19–21], 2D cameras [22–27], stereo cameras [28–37], etc. However, the detection performance of electromagnetic sensors is usually effected by the noisy signal emitted from other electric equipment. There are obvious shortcomings with the 2D detection, such as unavailable for measuring shapes, vulnerable to variable lighting conditions, and limited contrast compensation. Stereo cameras still need sufficient ambient light because of the structure; even the local feature matching was proposed in [31] to increase robustness to lighting.

To solve the problems mentioned above in object detection with 2D vision, the 3D point cloud technique has been developed to solve robust detection and precise localization. The point clouds of targeted objects can be acquired by 3D cameras to measure shapes and depth information. However, 3D point clouds exhibit the issues related to irregular distribution and structural ambiguity between the targeted objects and background. The object detection tasks of a mobile manipulator based on point clouds remain challenging because the performance can be influenced by background variations, irregular distribution, and structural ambiguity of point clouds. To suppress the above uncertain impact and extract features, deep learning algorithms can be utilized, such as grid-based methods: 3D Region Proposal Network (RPN) [38,39], 3D sparse convolution [40], and point-based methods: deep hough voting [41], Point Region Based Convolutional Neural Networks (PointRCNN) [42]. PointVoxel-Region Based Convolutional Neural Networks (PV-RCNN) takes the advantages of the grid-based and point-based techniques to solve the problem of object detection in a 3D environment and shows a good performance in 3D detection [43].

In this work, our main objective is to develop a collaborative industrial mobile manipulator integrated with an adaptive gripper and a learning-based computer vision system, providing mobility and manipulability and conducting detection and localization for production tasks in SMEs. To test the system, two exploratory tasks, including automating name tag production and plug-in charging, are investigated through the collaboration with our industrial partner, a Danish SME, Jydsk Emblem Fabrik A/S, referred to as JEF. It is a provider of name tags, medals, emblems, and uniform logos, or almost any possible objects which can be printed or engraved. As shown in Fig. 1, the mobile manipulator carrying a 3D camera, a 2D camera, and an adaptive gripper, needs to detect different workstations, navigate to the working positions of workstations, and localize the positions of printer button, handle of the lid based on calibration results. Precise manipulation in different workstations should be processed with the detection algorithm, calibration method, and adaptive gripper tools developed in this work.

The rest sections of this paper are organized as follows. Section 2 describes the problem statement and general solution framework of this project. Section 3 presents the technical details of the detection and localization method of the target objects. A point cloud dataset is also established in this section. Section 4 systematically introduces the system setup including the gripper design and main tools used. The detection and localization methods are validated through experimental methods in a SME factory in Section 5. The last section gives the conclusion and discussion.

2. Problem statement and general solution framework

The automation tasks in SMEs studied in this work include detecting, localizing the targeted objects, automating name tags production, and plug-in charging by robots. Detecting and localizing the targeted objects enable robots to obtain precise positions to navigate in dynamic environments. The production tasks can be completed based on detection, localization using the adaptive gripper precisely. The challenging issues include: (1) an image processing technique that can detect and localize the printing machine, laser cutter, charging station, plug, socket, and humans accurately; (2) the image processing technique which is robust against inconsistent light conditions and the uncertainty of point clouds; (3) the position of the mobile manipulator relative to the workspace; (4) the mobile manipulator which is adaptive and is able to automate the name tag production and plug-in charging tasks precisely.

Therefore, we employ an industrial mobile manipulator to automate production tasks due to its flexibility and larger workspace. To automate the tasks precisely, the industrial mobile manipulator we use is shown in Fig. 2, and mainly consists of a mobile platform (MiR 200), a collaborative manipulator (UR5e) [44], and a low-cost 2D camera and a 3D camera mounted at its end. The vacuum module is designed to pick up the big and flat wood sheets for producing name tags. Then, the 2D camera is used to calibrate the position of the mobile manipulator relative to the workspace. The 3D camera is employed to detect and localize targeted objects because it can acquire shapes and depth data. We establish a point cloud dataset containing the point clouds of the printing machine, laser cutter, charging station, plug, sockets, and humans by following KITTI Vision Benchmark [45]. The dataset is divided into the training dataset and the evaluation dataset. We utilize PV-RCNN to solve the uncertain issues of point clouds and increase the robustness of the detection. The performance of PV-RCNN in the 3D point cloud detection of autonomous driving has been demonstrated and proved with good performance. Therefore, it is employed to process 3D point clouds of detecting targeted objects. The whole detection procedure of automating production tasks and plug-in charging is shown in Fig. 3. The industrial mobile manipulator carrying two cameras and a gripper is utilized to detect and localize the printing machine, laser cutter, charging station, plug, socket, and

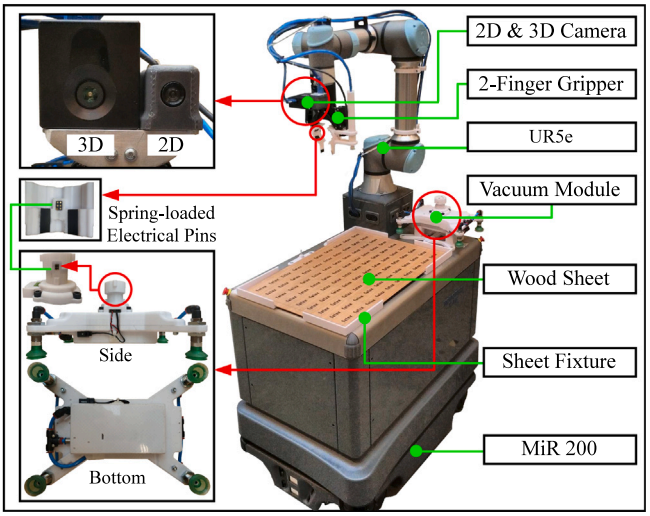

Fig. 2. The structure of the mobile manipulator using in SME.

humans using trained PV-RCNN. Then, the working positions relative to two machines can be calculated based on the detection results. For operating printing and cutting wood sheets precisely, the relative position between the mobile manipulator and the components of the machines are calibrated with chessboards using the 2D camera. The manipulator can then insert, remove and transport wood sheets, open and close the lid of the laser cutter based on the calibration results. For the self-charging task, the position and orientation of the socket and plug can be detected and calculated when the mobile manipulator arrives at the nearfield of the charging station. Finally, the coordinates of slot holes of the socket and grasping points of the plug serve as the input to the mobile manipulator to automate the plug-in charging task.

3. Detection and localization

In this section, we introduce the technical details of the detection and localization of the target objects. As with all deep learning models based on supervised learning, good ground truth for training is necessary to obtain a successful and robust object detector [46]. Therefore, the data acquisition and dataset establishment methods are introduced firstly. Then, we briefly present the principles and procedures of PV-RCNN used in the detection tasks. The localization of the mobile robot can be conducted using pre-established maps built by SLAM. However, it will generate localization errors when the layout of nearby targets is changed. Therefore, the precise localization is calculated by the transformation matrix from the targeted object to the robot based on detection results. Finally, the evaluation of the detection method is presented, and the performance of the detection method is evaluated through the experimental results.

3.1. Data acquisition

For point cloud acquisition, apart from the robustness to illumination mentioned above, the size and sampling frequency of cameras also need to be considered. For this reason, a 3D camera based on the Time of Flight (ToF) technique is employed, it can generate point clouds directly with a sampling rate of up to 60 fps. We take advantage of the mobile manipulator system equipped with a 3D camera to acquire point clouds. Therefore, in this research, we use a PMD 3D camera to acquire point clouds. For acquiring the point clouds of the printing machine, laser cutter, charging station, and humans, the height of the camera is set to be 91.3 cm above the ground and 79.6 cm for acquiring the point clouds of the socket and plug. The detection range is within [0, 9.6] m for the X-axis (depth), [-4.8, 4.8] m for the Y-axis (width), and [−2, 0] m for the Z-axis (height). The camera is mounted on the fixture of the gripper to acquire point clouds under ten fps acquiring frequency. Point clouds from different viewpoints, lighting conditions, and occlusion are acquired while the mobile manipulator navigates in the working environment to increase the robustness for training the predictive model of PV-RCNN.

3.2. Dataset

For supervised learning, the diversity and size of the dataset used to train the deep learning algorithm can affect the detection and localization accuracy. The inspiration for the detection and localization methods comes from the autonomous driving techniques in which objects need to be detected in real-time. Therefore, the setup standards of the dataset are referred to as the popular autonomous driving dataset-KITTI Vision Benchmark Suite [45]. The point cloud dataset consists of five parts: the point clouds of the printing machine, laser cutter, charging station, and humans, and the point clouds of the plug and socket. Note that the big difference in size between the plug, socket, and others, leads to the complexity issue of labeling the data and training performance. Thus, 10x is applied to the size of raw point clouds of sockets and plugs to keep the same ground truth level as KITTI without changing the origin center of the 3D camera. Several reasons for the above procedure are: (1) the plug and socket with bigger geometric size are easy to label; (2) the point clouds can keep the same ground truth level as KITTI dataset; (3) we can use the same training parameters for other target objects, and the socket and plug. After detecting processes, we restore the point clouds of the plugs and sockets to their original size to avoid influencing the position and localization. Note that there are 4400 training samples, which are divided into the training dataset (4000 samples) and the evaluation dataset (400 samples).

Finally, the raw 3D point clouds are labeled by cuboid annotation using the tool SUPERVISELY[47]. Each object is assigned with a class and its dimension (height, width, length). The position and orientation of each frame are also provided. The dataset and the tool we used can be found at https://github.com/Gltina/ACP-3Detection.

3.3. PV-RCNN

Firstly, the fundamental principles of PV-RCNN to process 3D point clouds are introduced. Then, 3D object detection procedures, which are included in Fig. 3, are illustrated. The advantages and the application of PV-RCNN on automatic production and plug-in charging tasks are introduced.

The fundamental principle of PV-RCNN is to integrate the point-based set abstraction and voxel-based sparse convolution into one network to learn more discriminative features of point clouds. The overall architecture of PV-RCNN mainly consists of two steps: the multi-scale semantic features are encoded from different layers, and then the features are aggregated to RoI grids for box proposal refinement, shown in Fig. 4. PV-RCNN takes the advantages of the grid-based and the point-based methods to learn more features of point clouds, and it is a well-established, open-source deep learning algorithm, and the adjustable voxelsize adapts to detect the objects in our research. Therefore, we train PV-RCNN using the datasets including the printing machine, laser cutter, charging station, plug, socket, and humans, socket and plug. We integrate the trained model into a mobile manipulator system for the detection of the above target objects. Two essential steps of PV-RCNN are introduced.

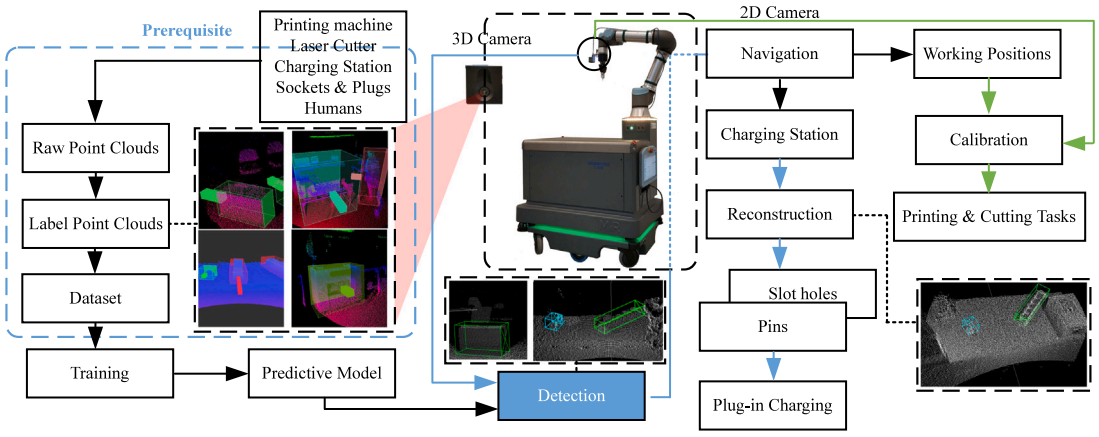

Fig. 3. The whole steps of automatic production and plug-in task in SME.

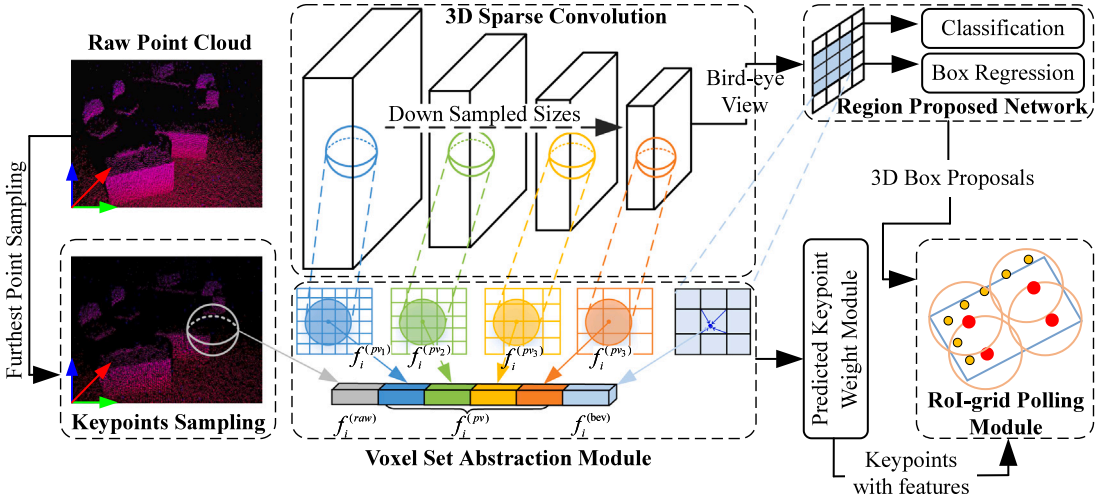

Fig. 4. The overall architecture of the PV-RCNN.

3.3.1. Voxel-to-keypoint scene encoding via voxel set abstraction

This step aggregates the multi-scale feature voxels into a set of keypoints. The multi-scale semantic feature for the key point p_i can be generated by aggregating the above features from different levels of the 3D voxel CNN.

$$f_i^{(pv)} = \left[f_i^{(pv_1)}, f_i^{(pv_2)}, f_i^{(pv_3)}, f_i^{(pv_4)} \right], i = 1, \ldots, n \tag{1}$$

where $f_i^{(pv_1)}$, $f_i^{(pv_2)}$, $f_i^{(pv_3)}$, $f_i^{(pv_4)}$ are the feature vectors from 4 layers respectively.

Then the keypoint features can be extended from the raw point clouds and the bird-view feature maps. Hence, the keypoint feature for p_i is further enriched as

$$f_i^{(p)} = \left[f_i^{(pv)}, f_i^{(raw)}, f_i^{(bev)} \right], i = 1, \ldots, n \tag{2}$$

where $f_i^{(raw)}$ is the raw point-cloud feature, $f_i^{(bev)}$ is the bird-view feature.

3.3.2. Keypoint-to-grid RoI feature abstraction

For accurate and robust proposal refinement, 3D proposal (RoI) features are aggregated from the keypoint features. Therefore, the keypoint-to-grid RoI feature abstraction is proposed based on the set abstraction operation for multi-scale RoI feature encoding. As shown in Fig. 5, the point cloud of the laser cutter is taken as a case, and the RoI-grid pooling module is proposed to aggregate the keypoint features from the last step to the RoI-grid points. Each 3D proposal

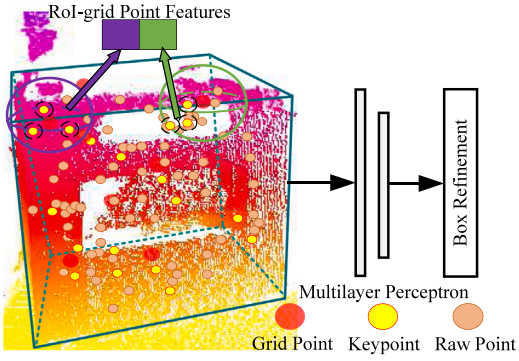

Fig. 5. Illustration of RoI-grid pooling module.

includes 6 × 6 × 6 grid points. The proposal refinement network is able to predict the size and location (i.e., center, size, and orientation) residuals relative to the input 3D proposal after obtaining each RoI feature. Box refinement can be achieved by the refinement network, which adopts a 2-layer multilayer perceptron (MLP), and the position, orientation, and dimension of cuboid boxes can be obtained.

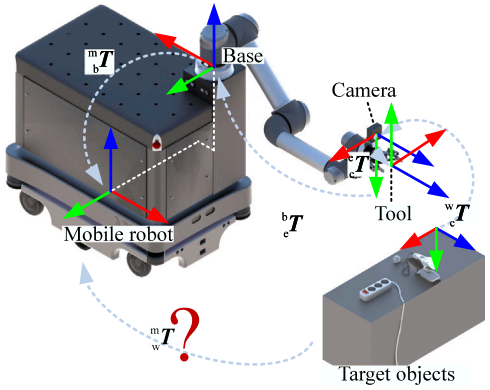

Fig. 6. Coordinates and transformation matrices of the system.

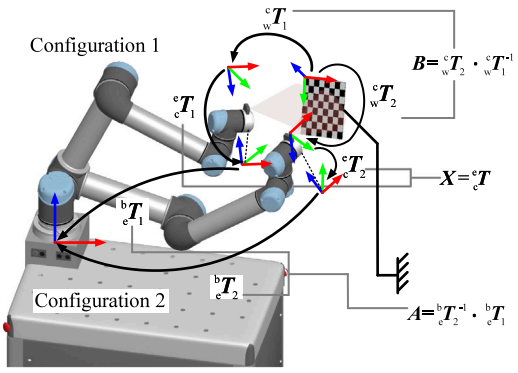

Fig. 7. Geometrical transformation for eye-in-hand calibration.

3.4. Localization

The predictive model of PV-RCNN can be used to detect the targeted objects in this paper. The relative position between the robot and the objects have to be obtained for navigating the robot among different working stations. As shown in Fig. 6, the relative position can be calculated by the transformation matrices from targeted objects to the mobile robot. The coordinate frames associated with the mobile robot, the base of the manipulator, tool center point (TCP) of the manipulator, cameras, and the targeted objects are indicated by the sub-indexes m, b, e, c, and w, respectively.

Therefore, the transformation matrix from the targeted objects to the mobile robot can be formulated as

$$ {}^{\mathrm{m}}_{\mathrm{w}}\boldsymbol{T} = {}^{\mathrm{m}}_{\mathrm{b}}\boldsymbol{T} \cdot {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T} \cdot {}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{T} \cdot {}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T} \tag{3} $$

where the matrix ${}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}$ denotes the transformation matrix from the targeted objects the 3D camera, the ${}^{\mathrm{m}}_{\mathrm{b}}\boldsymbol{T}$ can be identified by considering the mobile robot as the last joint of the manipulator, the parameters can be measured. Finally, the position of the targeted objects related to the mobile platform can be calculated by Eq. (3).

It is worth mentioning that the transformation matrix ${}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{T}$ from the camera to the end-effector can be obtained by the hand-eye calibration method, and the matrix ${}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}$ from the TCP to the base of the manipulator can be obtained by the kinematics of the manipulator.

3.4.1. Eye-in-hand calibration

As shown in Fig. 7, the formulation of the rigid transformations in the loop among camera, end-effector, and calibration board can be expressed as:

$$ {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_1 \cdot {}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{T}_1 \cdot {}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_1 = {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_2 \cdot {}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{T}_2 \cdot {}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_2 \tag{4} $$

Left multiplying both sides of the equation by ${}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_2^{-1}$, and right multiplying them by ${}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_1^{-1}$, we have

$$ {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_2^{-1} \cdot {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_1 \cdot {}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{T} = {}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{T} \cdot {}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_2 \cdot {}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_1^{-1} \tag{5} $$

${}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_2^{-1} \cdot {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_1$ can be interpreted as the movements made by the manipulator and denoted by:

$$ \boldsymbol{A} = {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_2^{-1} \cdot {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_1 = {}^{\mathrm{e}}_{\mathrm{b}}\boldsymbol{T}_2 \cdot {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}_1 = {}^{\mathrm{e2}}_{\mathrm{e1}}\boldsymbol{T} \tag{6} $$

Similarly, ${}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_2 \cdot {}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_1^{-1}$ can be denoted by $\boldsymbol{B}$ and interpreted as the movements of the camera:

$$ \boldsymbol{B} = {}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_2 \cdot {}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_1^{-1} = {}^{\mathrm{c}}_{\mathrm{w}}\boldsymbol{T}_2 \cdot {}^{\mathrm{w}}_{\mathrm{c}}\boldsymbol{T}_1 = {}^{\mathrm{c2}}_{\mathrm{c1}}\boldsymbol{T} \tag{7} $$

Then, Eq. (4) can be expressed by

$$ \boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{X} \cdot \boldsymbol{B} \tag{8} $$

Eq. (8) can be expanded and described as:

$$ \begin{bmatrix} {}^{ej}_{ei}\boldsymbol{R} & {}^{ej}_{ei}\boldsymbol{t} \\ 0_{1\times3} & 1 \end{bmatrix} \begin{bmatrix} {}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{R} & {}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{t} \\ 0_{1\times3} & 1 \end{bmatrix} = \begin{bmatrix} {}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{R} & {}^{\mathrm{e}}_{\mathrm{c}}\boldsymbol{t} \\ 0_{1\times3} & 1 \end{bmatrix} \begin{bmatrix} {}^{cj}_{ci}\boldsymbol{R} & {}^{cj}_{ci}\boldsymbol{t} \\ 0_{1\times3} & 1 \end{bmatrix} \tag{9} $$

where $\boldsymbol{R}$ is the rotation matrix, and $\boldsymbol{t}$ is translation vector.

Eqs. (8) and (9) represent the geometrical relationship between the coordinate frames of the camera, end-effector, and the manipulator base. In order to attain a solution for to Eq. (9) , at least three positions with non-parallel movements of the rotational axis are required [48]. If Eq. (8) has infinite solutions, the coefficients ($\boldsymbol{X}$) matrix should be singular. When high precision is required, more equations can be obtained after moving the manipulator to another position that is non-parallel to the previous one. Each of the equations is expressed as

$$ \boldsymbol{A}_i \cdot \boldsymbol{X} = \boldsymbol{X} \cdot \boldsymbol{B}_i, i = 1, \ldots, n \tag{10} $$

To solve the above equation, the rotation matrix, the rotation matrix ${}^{e}_{c}\boldsymbol{R}$ can be obtained using Rodrigues's Formula [48].

3.4.2. Kinematic model of the mobile manipulator

The transformation matrix from the end-effector of UR5e to the mobile robot is expressed as

$$ {}^{\mathrm{m}}_{\mathrm{e}}\boldsymbol{T} = {}^{\mathrm{m}}_{\mathrm{b}}\boldsymbol{T} \cdot {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T} \tag{11} $$

The transformation matrix ${}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}$ can be calculated by the forward kinematics of UR5e using the D-H parameters [49]. The forward kinematics model of the manipulator can be obtained by multiplying the six transformation matrices:

$$ {}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T} = {}^{0}_{6}\boldsymbol{T} = {}^{0}_{1}\boldsymbol{T} \cdot {}^{1}_{2}\boldsymbol{T} \cdot {}^{2}_{3}\boldsymbol{T} \cdot {}^{3}_{4}\boldsymbol{T} \cdot {}^{4}_{5}\boldsymbol{T} \cdot {}^{5}_{6}\boldsymbol{T} \tag{12} $$

The forward kinematics denotes the position and orientation of each link frame working forward from the base of the robot, and it can be used to obtain ${}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{T}$ in Eqs. (3) and (4).

The MiR 200 is a six-wheel robot with two drive wheels in the middle and four caster wheels to ensure stability when the platform rotates. The schematic diagram of MiR 200 in a world coordinate is shown in Fig. 8.

To specify the position of the MiR 200 on the world coordinate, the relationship between the two coordinates is established. The equations of the movement of MiR 200 are:

$$ \dot{X}_{\mathrm{M}} = (V_{\mathrm{L}} + V_{\mathrm{R}}) \cos \frac{\theta_{\mathrm{M}}}{2} \tag{13} $$

$$ \dot{Y}_{\mathrm{M}} = (V_{\mathrm{L}} + V_{\mathrm{R}}) \sin \frac{\theta_{\mathrm{M}}}{2} \tag{14} $$

$$ \dot{\theta}_{\mathrm{M}} = \frac{V_{\mathrm{L}} - V_{\mathrm{R}}}{2 \cdot L_1} \tag{15} $$

where the V_{L} and V_{R} are the linear velocity of the left and right wheels, respectively. $\dot{\theta}_{\mathrm{M}}$ is the angular velocity of the MiR 200, L_1 is the distance between two wheels.

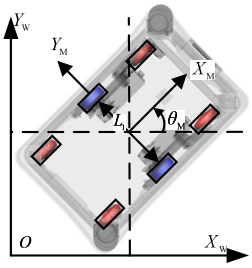

Fig. 8. The world coordinate and the Mir 200 local coordinate.

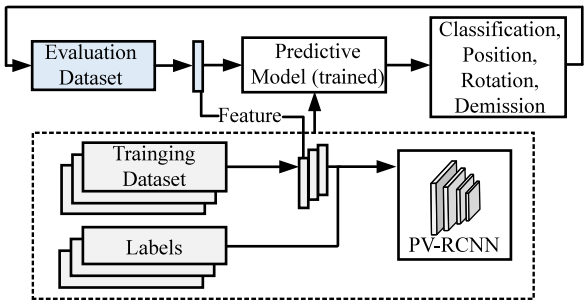

Fig. 9. The evaluation procedure of PV-RCNN.

Finally, the mobile robot can navigate to the working positions based on the position targeted objects relative to the base coordinate of the manipulator. Note that the mobile robot has 3-DOF, only x, y and Rz of targeted objects under base coordinate are utilized to calculate the relative position based on ${}^{\mathrm{m}}_{\mathrm{b}}\boldsymbol{T}$.

3.5. Testing of the detection

The process to test the model of PV-RCNN for detecting the printing machine, laser cutter, charging station, plug, socket and humans is shown in Fig. 9. The datasets consist of two parts: the training dataset (4000 samples) and the evaluation dataset (400 samples). To evaluate the trained model, we take 100 point clouds from the evaluation dataset one by one as the input to the PV-RCNN model trained by the point clouds of different target objects, respectively. The classification, position, orientation, dimension data of the cuboid box is obtained from the detection results. The detecting performance of the predictive model is evaluated by the overlap volumes between the cuboid boxes and ground truth boxes based on the evaluation dataset.

For training the PV-RCNN, the network and training details can be set as: the feature dimensions of the four-level 3D voxel CNN are 16, 32, 64, 64, respectively. The two neighboring radii of each grid point in the RoI-grid pooling operation are (0.8 m, 1.6 m). The voxel size is (0.03 m, 0.03 m, 0.05 m) in each axis. For our dataset, we train the network with the batch size 4, learning rate 0.01 for 1000 epochs on GEFORCE RTX 3090 GPU, which takes around 65 h for each targeted object.

Note that the number of cloud points in the training dataset and the epoch of CNN can play an essential role in the performance of the detection of the PV-RCNN model. Therefore, we investigate the effect of the two factors on the detection performance. We train the PV-RCNN using 100, 300, 600, 1000 point clouds under different epochs, which are from 20 to 1000. We calculate the average overlap volume between the cuboid boxes and ground truth boxes based on 100 evaluation point clouds. The evaluation results of the printing machine, laser cutter are shown in Figs. 10 and 11.

The figure shows the detection results are stable enough when the number of cloud points is 1000 and epochs are bigger than 500. The

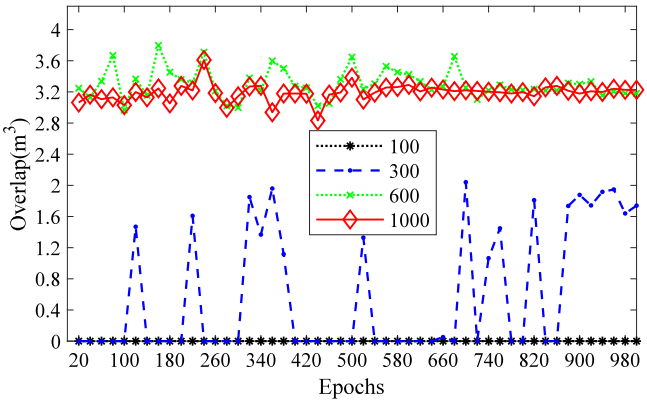

Fig. 10. The average overlap volumes of the printing machine.

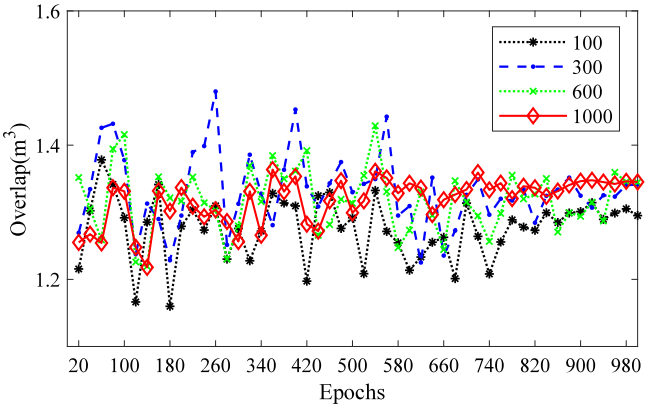

Fig. 11. The average overlap volumes of the laser cutter.

printing machine's detection results show worse performance when the dataset is less than 600 compared with the laser cutter because of its more complex structure. Under 1000 cloud points and 500 epochs, the detection success rates of the printing machine, laser cutter are 100% when there exist partial occlusions and variable light conditions. Therefore, we use 1000 point clouds to train the predictive model of the charging station, humans, plug and socket under the epochs from 20 to 500. For comparing the impact of cloud points on datasets, we also train the model using 300 point clouds. We calculate the average overlap volume using the evaluation dataset. The results of the charging station, humans, plug, and socket are shown in Figs. 12, 13, 14, respectively. As shown from Figs. 12 to 14, the detection results of the volumes of the targeted objects become steady when the epoch is bigger than 250. The evaluation method using overlap volume is inspired by the classical evaluation method Intersection-over-Union (IoU). The predictive models of the printing machine, laser cutter, charging station, humans, plug, and socket show good performance for detecting. The automatic name tags production and robotic plug-in charging experiment will be presented based on the detection and localization results in Section 5.

4. System setup

In order to use the mobile manipulator to perform tasks in the SME production, the UR5e collaborative industrial manipulator, Robotiq 2F-85 gripper, PMD 3D camera, See3Cam camera, and MiR200 mobile base form the main subsystems. A two-finger gripper based on Robotiq 2F-85 is designed and equipped on the end-effector of the manipulator, as shown in Fig. 2. The designed gripper is utilized to grasp the plug and the vacuum module, push the button of the printing machine and remove surplus material from the laser cutter. The vacuum module is

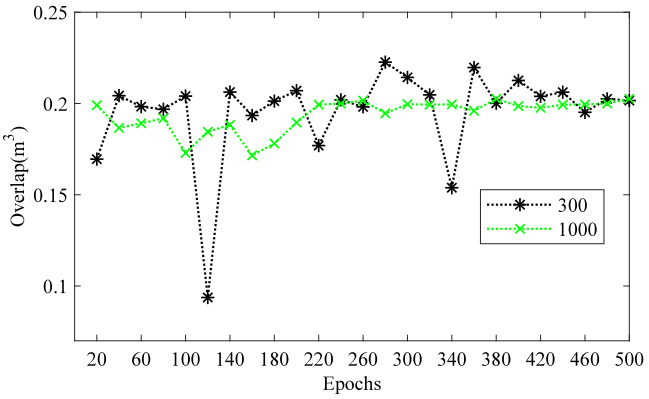

Fig. 12. The average overlap volumes of the charging station.

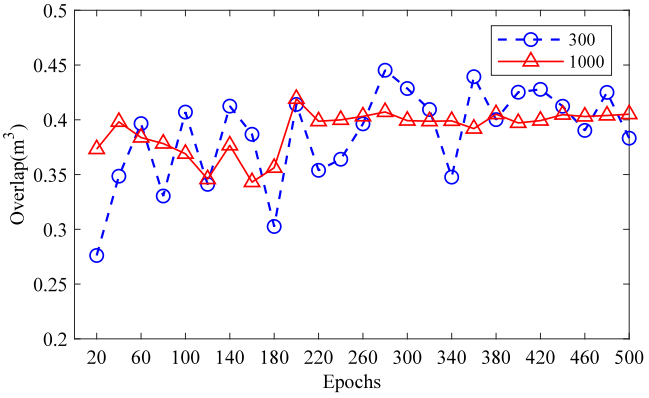

Fig. 13. The average overlap volumes of humans.

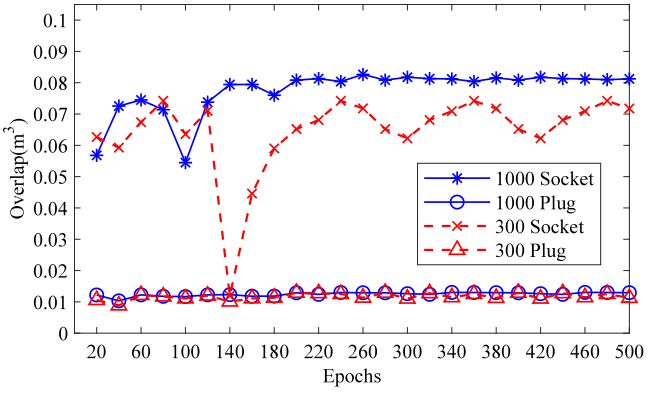

Fig. 14. The average overlap volumes of the socket and plug.

used to pick up the big and flat wood sheets; it connects the suction cups to the tool as a frame and simultaneously acts as an integrated container for the D2028 Vacuum pump and the FESTO MHE2-M1H-3/2O-M7 3/2 valve along with the pneumatic lines. See3Cam 2D camera and PMD 3D camera (CamBoard Pico Monstar) are mounted on the fixture of the gripper. See3Cam 2D camera has a resolution of 1920 × 1080. The 3D camera has wide Field of view (FoV) of 100° x 85° and a high pixel count of 352 x 288 pixels, and covers a massive 6 m measurement range.

4.1. Gripper design

The two-finger gripper is redesigned based on the commercial Robotiq 2F-85 gripper, which is mounted directly on the end-effector. It is a

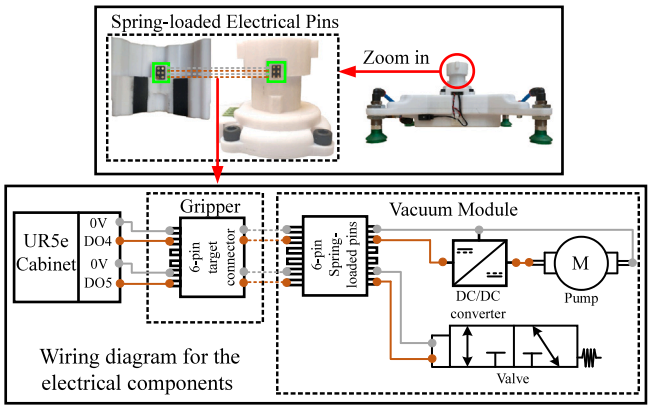

Fig. 15. The vacuum module and the wiring diagram of the electrical components.

multi-function gripper, which can be used to grasp the plug, the wood sheet after cutting and the function can be extended by connecting with the vacuum module in this research. It also can be used to pick up other production components in JEF, such as needles, name tags or medals for future research.

The two-finger gripper cannot pick up big and thin board-shaped components such as wood sheets and cardboards. This can be solved in several ways, where one option is to buy a tool changer such as [50]. A second option is to utilize the module, which can be grasped and charged by the two-finger gripper directly. Comparing the two options is quickly seen that the latter option has lower cycling time between tool changes, requires fewer total parts, and is more cost-beneficial. Therefore, a vacuum module is designed with the necessary geometry and electrical connections to power the on-board 12 V vacuum pump and the 24 V 3/2 valve. It consists of four Ø25 mm suction cups distanced by length and width of 238 mm and 166 mm, respectively. To ensure a stable electrical connection, spring-loaded pins are used, and the wiring is done internally, the wiring diagram for the electrical components is shown in Fig. 15.

4.2. Other tools

For automating production tasks by robots, several devices need to be designed and mounted on the machines. Precise calibration of the robot position relative to the workspace is required. To calculate the precise relative positions between the machines and the mobile manipulator, chessboards are mounted on the printing machine and laser cutter, respectively. The boards are used to calibrate the relative positions of the mobile manipulator to machines after arriving at the working positions. To open and close the lid of the laser cutter, a handle is designed and mounted onto the lid. The handle can rotate freely around an internal solid cylinder when grasped to increase the tolerance of opening the lid.

5. On-site testing

The primary purpose of our experiments is to validate the proposed 3D objection detection and localization with application to the automatic name tag production, automatic plug-in charging using the mobile manipulator in SMEs. The detection of humans can increase safety while interacting with humans. The predictive PV-RCNN for different target objects can be obtained using the method in Section 3. The flowchart of the experiment is shown in Fig. 16. The detection targets include the printing machine, laser cutter, charging station and humans. Firstly, the predictive model is employed to detect the printing machine, laser cutter and avoid humans. The model is utilized to detect the charging station, plug, and socket when the battery is low. Then the

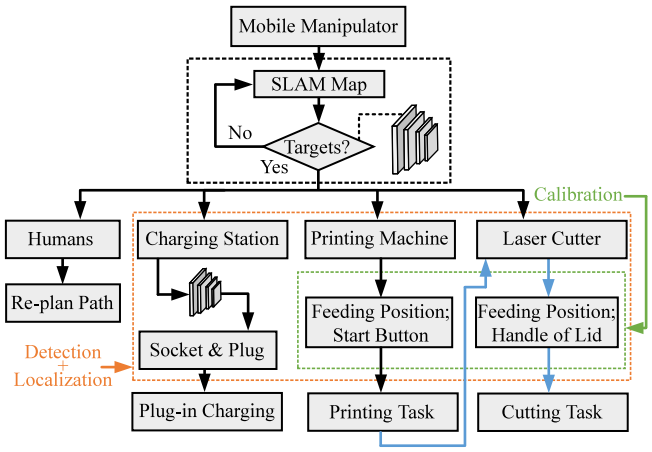

Fig. 16. Experiment flowchart.

experiment testing is performed in JEF; the target objects may not be in the field of view initially. Therefore, the mobile manipulator first navigates to the nearfield of the working stations based on the SLAM map. When the printing machine is detected through the predictive model, the mobile manipulator navigates to the working position based on the localization results. After detecting the target object and arriving at the working position, we employ a 2D camera to calibrate the robot position relative to the printing machine. The precise relative positions between the mobile manipulator and components of the machines such as, the button of the printer and the inserting position can be calculated. The wood sheet is picked up and placed at the inserting position obtained based on calibration results. When the printing is finished, the wood sheet is transported to the laser cutter using the same detecting and localizing strategies. Simultaneously, the mobile manipulator re-plans the path or halt when humans are detected on its path. The predictive PV-RCNN model can also be utilized to detect the charging station when the battery is low. After detecting the socket and plug, we reconstruct the point cloud of the charging station's surface to obtain the complete point cloud for calculating the position of pins and slot holes. The specific experimental tasks, including name tags manufacturing and plug-in charging, are also introduced in this section. Finally, we examine the success rate of the automatic printing and cutting of name tags and automatic plug-in charging by the mobile manipulator and analyze the results.

5.1. Name tag production

After training the PV-RCNN model in Section 3 by the datasets including the printing machine and laser cutter, it is integrated into the mobile manipulator for detecting the targeted objects for name tags manufacturing. The mobile manipulator navigates to the nearfield of the printing machine or laser cutter based on the SLAM results. However, the repeated movement of the mobile manipulator can generate the increasing error, which causes the chessboard to be beyond the camera's line of sight. Therefore, the working positions can be updated after detecting targeted objects. Before navigating, the mobile manipulator moves the 3D camera to 91.3 cm above the ground. The mobile robot can navigate in the testing environment, and the manipulator moves the end-effector from 0° to 180° to search the target objects while the mobile manipulator is moving to the working positions. When the target objects are detected, the mobile manipulator can navigate them based on the detection results. The detection rate is set as 24 fps. The relative positions between the mobile manipulator and the target objects are calculated by coordinate transformation matrices in Section 3.

The process of name tags production can be divided into two steps: printing and cutting the wood sheet. The whole piece of the wood sheet can be printed with pre-set pictures by printing machine firstly; the printed wood sheet can be delivered to be cut into several small pieces by the laser cutter. After the mobile manipulator arrives at a working position which is 20 cm away from a working station, the 2D camera is moved to the top of the chessboard mounted on the surface of the printing machine or laser cutter. The relative positions between the manipulator and components of the two machines can be calculated. In this subsection, the automatic procedures of printing and cutting are introduced in detail.

5.1.1. Automatic printing task

The process of automatic printing task described below is depicted in Fig. 17. After calibrating with the chessboard, the mobile manipulator moves the two-finger gripper to the position to grasp the vacuum module from the fixture. The combined gripper moves to the top of the wood sheet and picks it up. The position, where we insert the wood sheet into the printing machine, is calculated by referring to the chessboard. After placing the wood sheet on the vacuum screen, which contains suction holes on its surface and a vacuum pump inside, the wood sheet is pushed against alignment the fixture corner on the vacuum screen to increase the tolerance shown in step 5 in Fig. 17. Finally, the two-finger gripper drops the vacuum module to its fixture and pushes the running button on the printing machine.

The vacuum screen of the printing machine is to keep the wood sheet flat to conduct good printing performance. It is observed that the force to remove the wood sheet from the printing machine increases because of the air pressure from the vacuum screen. To investigate the effect of picking position to the success rate, an experiment has been carried out. Two picking positions, which are top left corner and center, have been selected, and the success rate for picking up the wood sheet from the vacuum screen is tested. The picking success rates are 100% and 0%. Therefore, the position at the top left corner is chosen as the picking position for the vacuum module.

After printing, the two-finger gripper grasps the vacuum module again and picks up the wood sheet from the printing machine. Then, the printed wood sheet is placed into the sheet fixture. The vacuum module is placed back to the fixture on the top of the mobile robot. Finally, the automatic printing task process is finished, and the printed wood sheet is prepared to be delivered to the laser cutter.

5.1.2. Automatic cutting task

The procedure of the automatic printing task is schematically shown in Fig. 18. When the mobile manipulator arrives at the working position for the laser cutter and calibrates with the chessboard, the mobile manipulator grasps the handle to open the lid of the laser cutter. Then, the surplus material is removed using the two-finger gripper. The vacuum module is grasped again to pick up the wood sheet to insert into the fixture of the laser cutter. The vacuum module is placed back to its fixture. The lid of the laser cutter is closed by the mobile manipulator and the machine is operated through sending the signal from the mobile manipulator to the PC which is used to control the laser cutter. The cut name tags fall into the bucket through a slope.

5.2. Automatic plug-in charging task

There are three main steps for the automatic plug-in charging task. After detecting the charging station, the relative position between the mobile manipulator and the charging station is calculated by the coordinate transformation matrix. The mobile manipulator navigates to 20 cm from the long side of the charging station. The 3D camera is then moved down to 79.6 cm above the ground to detect the socket and plug. After detection, the point cloud is sparse so that dense point cloud is necessary to obtain the accurate position of the pins of the plug and the slot holes of the socket. Therefore, we develop a 3D reconstruction

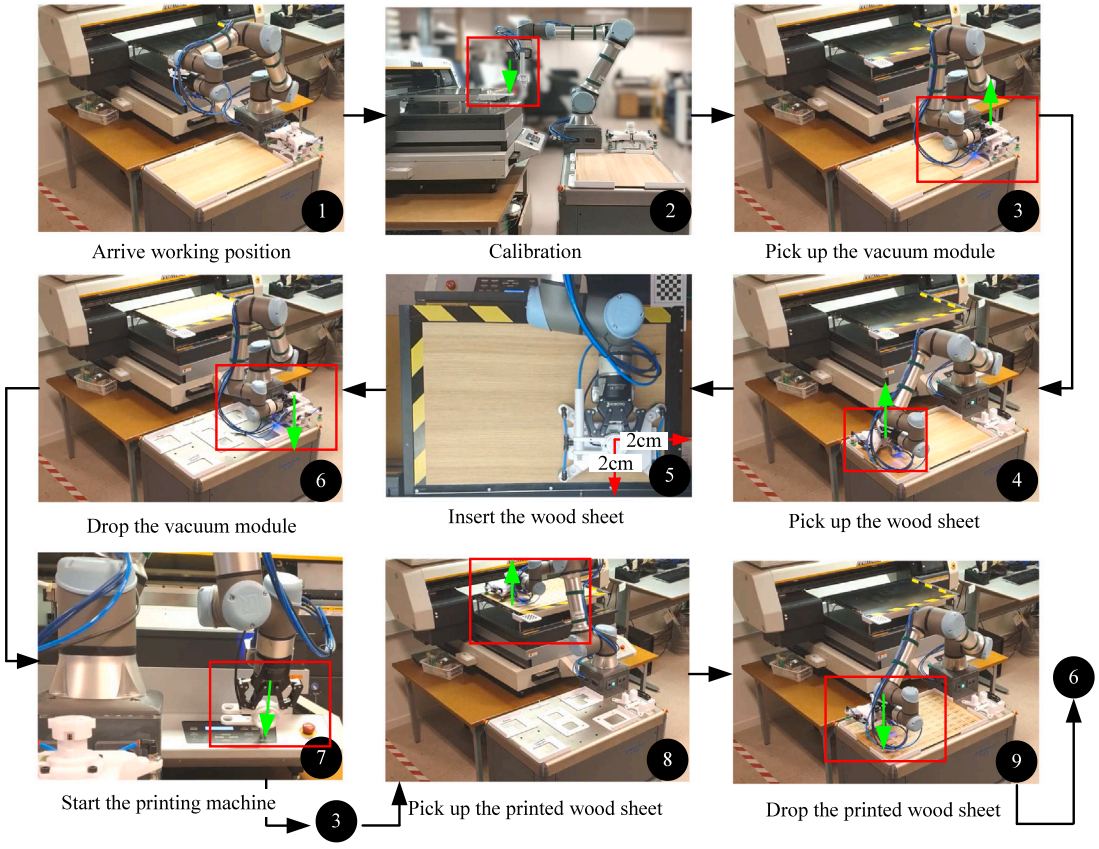

Fig. 17. Process of the automatic wood sheet printing.

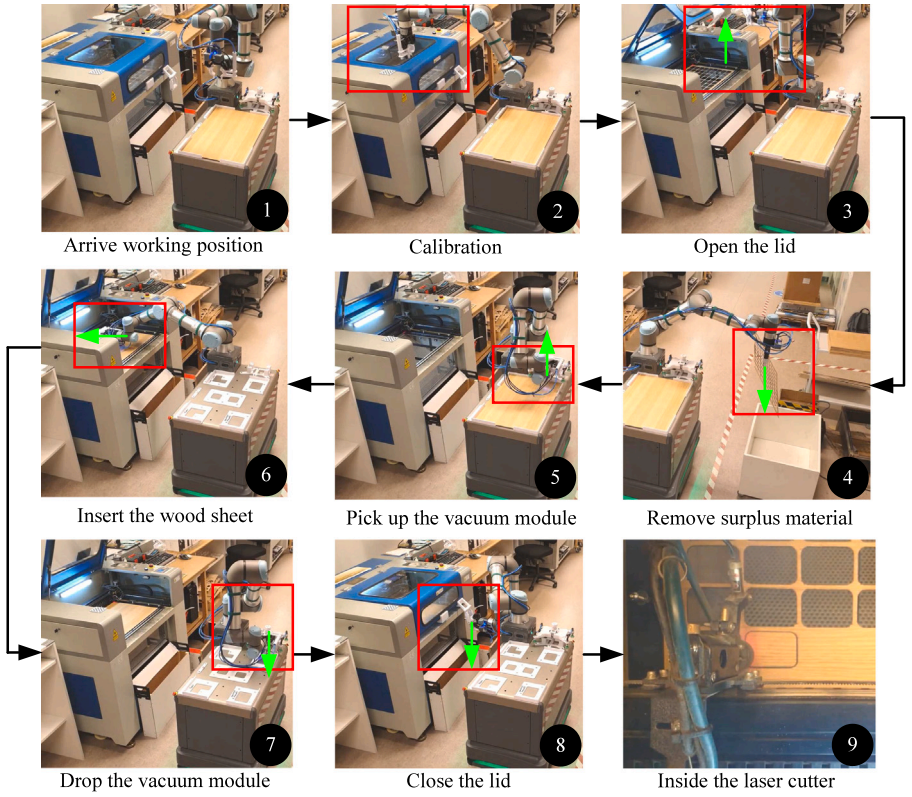

Fig. 18. Process of the automatic wood sheet cutting.

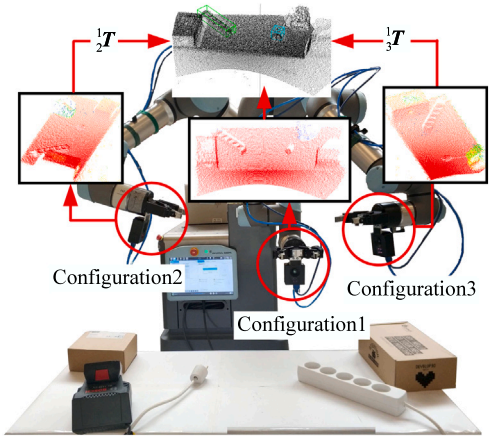

Fig. 19. Robot experimental testing of automatic plug-in charging.

method using the mobile manipulator. The experiment is shown in Fig. 19.

We acquire other two point clouds from two different positions where are 30 cm left and right from the current position of the end-effector. The three cloud points and corresponding configurations are acquired simultaneously. The reconstruction of cloud points is obtained based on the transformation matrices among different configurations. In this research, we mainly focus on the detection task. We select the grasping point positions of the plug and the slot holes position of the socket from a point cloud. The manipulator can grasp and insert the plug into the socket using the positions and orientations of the plug and socket.

Safety is a primary prerequisite for mobile robot manipulation when humans work alongside mobile manipulators. Specifically, to achieve the safety in SMEs environment, the collision avoidance, collision reaction, and tip-over are several main challenges for mobile manipulators. The collision between robots and surrounding humans or objects could be avoided by maintaining safe operational distances through the external position monitoring sensors, such as security laser scanners [51] or safe camera systems [52]. Thus, the mobile manipulator can operate at different speeds depending on different safe distances. The potential solution to reducing collision risk is removing power to the drives, applying brakes, and actively controlling motors to counter motion through control systems based on electric current and sensor systems, such as, force and torque sensors in joints and tactile sensors over robots, etc [53]. The tip-over stability of mobile manipulators is effected by slopes of the ground or their dynamics. Potential solutions to tip-over issues include FA measuring method [54], adaptive neural fuzzy approach [55], stability margin metric-increment function [56], etc. However, the detailed safety solutions of mobile robot manipulators are beyond the scope of this article, but they could be our future efforts.

5.3. Testing results

Since the mobile manipulator aligns the top left corner of the sheet against the fixture in the printer shown in step 5 of Fig. 17, the measuring of the sheet is done with respect to that corner. While measuring, the sheet is fixed against a similar corner to make it easier to do the measurements. The test pattern is 90 rectangles with a text inside. Out of those 90 rectangles, the four corner rectangles marked as 1, 2, 3, and 4 in Fig. 20 are measured to explain the error of the process. The expected (X, Y)-coordinates of the 4 rectangles measured from the top left corner of the sheet are: $(X_1,Y_1)_{expected}$ = (32.5 mm, 23 mm), $(X_2,Y_2)_{expected}$ = (527.5 mm, 23 mm), $(X_3,Y_3)_{expected}$ = (32.5 mm, 351 mm) and $(X_4,Y_4)_{expected}$ = (527.5 mm, 351 mm).

Fig. 20. Measuring of printed sheet with test pattern.

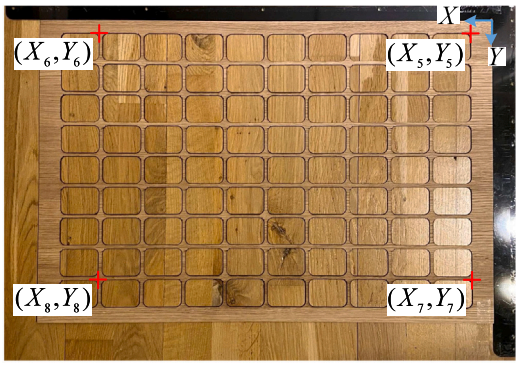

Fig. 21. Measuring of laser cut sheet with test pattern.

The sheet is measured using a ruler with a resolution of 0.5 mm. The reason why the pattern is only measured against the corner is that the shorter sides of the sheets are not perfectly straight, and also the sheets are not exactly 600 mm × 400 mm. Therefore the only known reference point for the pattern on the sheet is the corner that was aligned with the fixture during printing. After printing ten wood sheets and mearing ten sets of data, most of the measurements are close to the expected values. The most significant error is 1 mm. The four lines created by the four points are used to calculate the average rotation error, which is 0.3°.

Since the mobile manipulator aligns the top right corner of the sheet against the fixture in the laser cutter, the measuring of the sheet is done with respect to that corner of the sheet. The test pattern is, as mentioned 90 rectangular tags, resulting in 90 holes. Out of those 90, the four corner holes marked as 5, 6, 7, and 8 in Fig. 21 are measured to determine the error of the process. The expected (X,Y)-coordinates measured from the top right corner of the sheet are: (X_5, Y_5) = (27.5 mm, 18 mm), (X_6, Y_6) = (522.5 mm, 18 mm), (X_7, Y_7) = (27.5 mm, 346 mm) and (X_8, Y_8) = (522.5 mm, 346 mm). The measurement is the same as the method for printing. After cutting ten wood sheets and measuring ten sets of data, most of the measurements are close to the expected values. The biggest error is 1.5 mm. The four lines created by the four points are used to calculate the average rotation error, which is 0.3°.

Based on these ten sheets with 900 tags, it seems like the maximum error in Y is 1 mm and no measurable error is found in the X direction for printing task, and maximum error in Y is 1.5 mm and in X is 0.5 mm for cutting task. The angular error is all about 0.3°. The maximum translational error is simply 1 mm. The reason for the errors could be the fact that the fixture position in the printer has not been calibrated. It could also be that the mobile manipulator did not push the sheet correctly against the corner. Lastly, the measuring process might have

Table 1
The requirements and performance of the automatic production tasks.

Tasks	Requirements		
	Automation (Status)	Cycle time (Status)	Functions (Status)
Name tags printing	Fully automatic process (✓)	2.4 s/tag (✓)	1. Start the printer (✓); 2. Adaptive to other printing tasks (✓); 3. Move the wood sheet to the cutter (✓).
Name tags cutting	Fully automatic process (✓)	1 min/sheet(**x**)	1. Start the cutter (✓); 2. Open and close the lid (✓).

influenced the result a bit since the resolution of the ruler is 0.5 mm and the measuring is done by hand.

The functional requirements and the performance of the mobile manipulator from JEF are shown in Table 1. The total time to insert and remove sheets from the printer is 1 min 09 s. The three name tag designs considered in this project yield 70–90 name tags, so the requirement is fulfilled. The total time to remove surplus material from the laser cutter and insert a new sheet is 1 min 39 s, while the requirement is 1 min. The requirements of a fully automated printing and cutting process have been achieved, and the robot can start the printing machine, laser cutter, and open or close the lid of the laser cutter.

Finally, for producing name tags by 10 wood sheets, the detection success rate of the printing machine, laser cutter and humans are 100%. We test the plug-in charging task using the mobile manipulator in JEF by 50 trials, the detection success rate of the charging station, the plug and socket are 100%. After navigating to the charging station and inputting the positions and orientations of the plug and socket, the plug-in success rate is 92%. The results validate the feasibility of the robotic solution framework for automating plug-in charging. It should be noted that the errors might come from the system error or the transformation matrix error.

6. Conclusion and discussion

We present a general robotic solution to automatic name tag production and plug-in charging in SMEs using the mobile manipulator, which is equipped with cameras and an adaptive gripper. We implement 3D point cloud technology to establish the datasets, including the training and evaluation parts of the detecting targets. In this solution, the detection of targeted objects has been obtained using the deep learning based the PV-RCNN and the localization has been performed based on the detection results. The 2D camera is employed to calibrate the relative position of the mobile manipulator to workstations. Our experimental results show the overlap volume as the performance index of the predictive deep learning model is accurate enough to automate name tags production and plug-in charging by the mobile manipulator. Finally, to validate the proposed framework, we utilize the mobile manipulator to automate two SME production tasks. The results of the robotic onsite experiments demonstrate that our method has a high detection accuracy at various light conditions and a high plug-in success rate.

The automatic solution and the computer vision system proposed in this paper represent a flexible and robust method for SMEs manufacturing. However, the performance of the 3D detection algorithm was not validated for small-size components in SMEs comparing with other literatures. The first important requirement of SME production is to ensure human's safety. The collision avoidance and human–robot physical interaction can be achieved by utilizing scanners or vision systems to avoid humans and following safety standards, such as, ISO 12100, IEC645508 [57]. Moreover, another important requirement is to reduce the manufacturing cycle time of mobile manipulators under the premise of safety as short as possible. Further improvements of the cycle time include optimizing the production process in SMEs or the motion planning of mobile manipulators. These two main requirements can further be achieved in our future efforts. We also plan to investigate the method for evaluating the performance of the detection algorithm considering the position and orientation errors between the cuboid boxes and the ground truth boxes for future work. Furthermore, applying mobile manipulators to more production tasks in SMEs is also a potential area for further research.

CRediT authorship contribution statement

Zhengxue Zhou: Writing – original draft, Methodology. **Leihui Li:** Software, 3D vision. **Alexander Fürsterling:** Hardware development. **Hjalte Joshua Durocher:** Software, Computer vision. **Jesper Mouridsen:** Time management, Neural Network. **Xuping Zhang:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by the European Union Regional Fund: Integrating Human-Robot Collaboration into Danish SME Manufacturing and Production based on the collaboration between Jydsk Emblem Fabrik A/S, NIZE Equipment ApS and the Department of Engineering at Aarhus University. Thanks for the support of the China Scholarship Council (NO. 201906080023) for my PhD study. More details of the project can be found in the website [58].

References

[1] V. Gopinath, K. Johansen, M. Derelöv, A.k. Gustafsson, S. Axelsson, Safe collaborative assembly on a continuously moving line with large industrial robots, Robot. Comput.-Integr. Manuf. 67 (2021) 102048.
[2] L. Gualtieri, E. Rauch, R. Vidoni, Emerging research fields in safety and ergonomics in industrial collaborative robotics: A systematic literature review, Robot. Comput.-Integr. Manuf. 67 (2021) 101998.
[3] A. Downs, Z. Kootbally, W. Harrison, P. Pilliptchak, B. Antonishek, M. Aksu, C. Schlenoff, S.K. Gupta, Assessing Industrial Robot agility through international competitions, Robot. Comput.-Integr. Manuf. 70 (2021) 102113.
[4] J.-S. Hyun, M.G. Carmichael, A. Tran, S. Zhang, D. Liu, Evaluation of fast, high-detail projected light 3D sensing for robots in construction, in: 2019 14th IEEE Conference on Industrial Electronics and Applications, ICIEA, IEEE, 2019, pp. 1262–1267.
[5] G. Michalos, S. Makris, P. Aivaliotis, S. Matthaiakis, A. Sardelis, G. Chryssolouris, Autonomous production systems using open architectures and mobile robotic structures, Proc. Cirp 28 (2015) 119–124.
[6] K. Zhou, G. Ebenhofer, C. Eitzinger, U. Zimmermann, C. Walter, J. Saenz, L.P. Castaño, M.A.F. Hernández, J.N. Oriol, Mobile manipulator is coming to aerospace manufacturing industry, in: 2014 IEEE International Symposium on Robotic and Sensors Environments (ROSE) Proceedings, IEEE, 2014, pp. 94–99.
[7] R.G. Rivera, R. García Alvarado, A. Martínez-Rocamora, F. Auat Cheein, A comprehensive performance evaluation of different mobile manipulators used as displaceable 3D printers of building elements for the construction industry, Sustainability 12 (11) (2020) 4378.
[8] A. Stopp, T. Baldauf, R. Hantsche, S. Horstmann, S. Kristensen, F. Lohnert, C. Priem, B. Ruscher, The manufacturing assistant: Safe, interactive teaching of operation sequences, in: Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication, IEEE, 2002, pp. 386–391.

[9] R. Jamisola, M.H. Ang, D. Oetomo, O. Khatib, T.M. Lim, S.Y. Lim, The operational space formulation implementation to aircraft canopy polishing using a mobile manipulator, in: Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292), Vol. 1, IEEE, 2002, pp. 400–405.

[10] B. Hamner, S. Koterba, J. Shi, R. Simmons, S. Singh, An autonomous mobile manipulator for assembly tasks, Auton. Robots 28 (1) (2010) 131–149.

[11] S. Datta, R. Ray, D. Banerji, Development of autonomous mobile robot with manipulator for manufacturing environment, Int. J. Adv. Manuf. Technol. 38 (5) (2008) 536–542.

[12] E. Helms, R.D. Schraft, M. Hagele, rob@ work: Robot assistant in industrial environments, in: Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication, IEEE, 2002, pp. 399–404.

[13] M. Hvilshøj, S. Bøgh, O. Madsen, M. Kristiansen, The mobile robot "Little Helper": concepts, ideas and working principles, in: 2009 IEEE Conference on Emerging Technologies & Factory Automation, IEEE, 2009, pp. 1–4.

[14] M. Hvilshøj, S. Bøgh, O. Madsen, M. Kristiansen, Calibration techniques for industrial mobile manipulators: Theoretical configurations and best practices, in: ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics), VDE, 2010, pp. 1–7.

[15] R.S. Andersen, J.S. Damgaard, O. Madsen, T.B. Moeslund, Fast calibration of industrial mobile robots to workstations using QR codes, in: IEEE ISR 2013, IEEE, 2013, pp. 1–6.

[16] O. Madsen, S. Bøgh, C. Schou, R.S. Andersen, J.S. Damgaard, M.R. Pedersen, V. Krüger, Integration of mobile manipulators in an industrial production, Ind. Robot: Int. J. (2015).

[17] M. Hvilshøj, S. Bøgh, O.S. Nielsen, O. Madsen, Multiple part feeding–real-world application for mobile manipulators, Assembly Autom. (2012).

[18] J. Aulinas, Y.R. Petillot, J. Salvi, X. Lladó, The slam problem: a survey, CCIA 184 (1) (2008) 363–371.

[19] B. Mayton, L. LeGrand, J.R. Smith, Robot, feed thyself: Plugging in to unmodified electrical outlets by sensing emitted ac electric fields, in: 2010 IEEE International Conference on Robotics and Automation, IEEE, 2010, pp. 715–722.

[20] T.D. Nguyen, T. Kim, J. Noh, H. Phung, G. Kang, H.R. Choi, Skin-type proximity sensor by using the change of electromagnetic field, IEEE Trans. Ind. Electron. 68 (3) (2020) 2379–2388.

[21] S. Song, Z. Li, H. Yu, H. Ren, Shape reconstruction for wire-driven flexible robots based on Bézier curve and electromagnetic positioning, Mechatronics 29 (2015) 28–35.

[22] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, P. Abbeel, Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding, in: 2010 IEEE International Conference on Robotics and Automation, IEEE, 2010, pp. 2308–2315.

[23] Z. Zheng, Y. Ma, H. Zheng, Y. Gu, M. Lin, Industrial part localization and grasping using a robotic arm guided by 2D monocular vision, Industrial Robot: An International Journal (2018).

[24] O. Kroeger, F. Wollschläger, J. Krüger, Low-cost embedded vision for industrial robots: A modular end-of-arm concept, in: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation, Vol. 1, ETFA, IEEE, 2020, pp. 1301–1304.

[25] C. Wang, L. Yin, Q. Zhao, W. Wang, C. Li, B. Luo, An intelligent robot for indoor substation inspection, Ind. Robot: Int. J. Robot. Res. Appl. (2020).

[26] A.S. Olesen, B.B. Gergaly, E.A. Ryberg, M.R. Thomsen, D. Chrysostomou, A collaborative robot cell for random bin-picking based on deep learning policies and a multi-gripper switching strategy, Proc. Manuf. 51 (2020) 3–10.

[27] A. Sardelis, N.-C. Zacharaki, Z. Arkouli, D. Andronas, G. Michalos, S. Makris, G. Papanikolopoulos, 2-Stage vision system for robotic handling of flexible objects, Proc. CIRP 97 (2021) 491–496.

[28] T.H. Luu, T.H. Tran, 3D vision for mobile robot manipulator on detecting and tracking target, in: 2015 15th International Conference on Control, Automation and Systems, ICCAS, IEEE, 2015, pp. 1560–1565.

[29] J. Kuehnle, A. Verl, Z. Xue, S. Ruehl, J.M. Zoellner, R. Dillmann, T. Grundmann, R. Eidenberger, R.D. Zoellner, 6d object localization and obstacle detection for collision-free manipulation with a mobile service robot, in: 2009 International Conference on Advanced Robotics, IEEE, 2009, pp. 1–6.

[30] W. Meeussen, M. Wise, S. Glaser, S. Chitta, C. McGann, P. Mihelich, E. Marder-Eppstein, M. Muja, V. Eruhimov, T. Foote, et al., Autonomous door opening and plugging in with a personal robot, in: 2010 IEEE International Conference on Robotics and Automation, IEEE, 2010, pp. 729–736.

[31] V. Eruhimov, W. Meeussen, Outlet detection and pose estimation for robot continuous operation, in: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2011, pp. 2941–2946.

[32] M. Dinham, G. Fang, Autonomous weld seam identification and localisation using eye-in-hand stereo vision for robotic arc welding, Robot. Comput.-Integr. Manuf. 29 (5) (2013) 288–301.

[33] J.-K. Oh, S. Lee, C.-H. Lee, Stereo vision based automation for a bin-picking solution, Int. J. Control Autom. Syst. 10 (2) (2012) 362–373.

[34] Y. Tang, L. Li, C. Wang, M. Chen, W. Feng, X. Zou, K. Huang, Real-time detection of surface deformation and strain in recycled aggregate concrete-filled steel tubular columns via four-ocular vision, Robot. Comput.-Integr. Manuf. 59 (2019) 36–46.

[35] G. Michalos, S. Makris, A. Eytan, S. Matthaiakis, G. Chryssolouris, Robot path correction using stereo vision system, Proc. CIRP 3 (2012) 352–357.

[36] S. Makris, Vision guided robots. Calibration and motion correction, in: Cooperating Robots for Flexible Manufacturing, Springer, 2021, pp. 205–222.

[37] L. Sabattini, A. Levratti, F. Venturi, E. Amplo, C. Fantuzzi, C. Secchi, Experimental comparison of 3D vision sensors for mobile robot localization for industrial application: Stereo-camera and RGB-D sensor, in: 2012 12th International Conference on Control Automation Robotics & Vision, ICARCV, IEEE, 2012, pp. 823–828.

[38] S. Song, J. Xiao, Deep sliding shapes for amodal 3d object detection in rgb-d images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 808–816.

[39] Y. Zhou, O. Tuzel, Voxelnet: End-to-end learning for point cloud based 3d object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4490–4499.

[40] B. Graham, M. Engelcke, L. Van Der Maaten, 3d semantic segmentation with submanifold sparse convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9224–9232.

[41] C.R. Qi, O. Litany, K. He, L.J. Guibas, Deep hough voting for 3d object detection in point clouds, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9277–9286.

[42] S. Shi, X. Wang, H. Li, Pointrcnn: 3d object proposal generation and detection from point cloud, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 770–779.

[43] S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, H. Li, Pv-rcnn: Point-voxel feature set abstraction for 3d object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10529–10538.

[44] Z. Zhou, L. Li, R. Wang, X. Zhang, Experimental eye-in-hand calibration for industrial mobile manipulators, in: 2020 IEEE International Conference on Mechatronics and Automation, ICMA, IEEE, 2020, pp. 582–587.

[45] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? the kitti vision benchmark suite, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 3354–3361.

[46] C. Li, Y. Huang, H. Li, X. Zhang, A weak supervision machine vision detection method based on artificial defect simulation, Knowl.-Based Syst. 208 (2020) 106466.

[47] SUPERVISE, https://supervise.ly/.

[48] R.Y. Tsai, R.K. Lenz, et al., A new technique for fully autonomous and efficient 3 D robotics hand/eye calibration, IEEE Trans. Robot. Autom. 5 (3) (1989) 345–358.

[49] Universal Robots, Parameters for calculations of kinematics and dynamics, https://www.universal-robots.com/articles/ur/parameters-for-calculations-of-kinematics-and-dynamics/.

[50] Schunk, Change systems, https://schunk.com/ru_en/gripping-systems/category/gripping-systems/handling/changing/change-systems/.

[51] H. Baltzakis, A. Argyros, P. Trahanias, Fusion of laser and visual data for robot motion planning and collision avoidance, Mach. Vis. Appl. 15 (2) (2003) 92–100.

[52] F. Flacco, T. Kroeger, A. De Luca, O. Khatib, A depth space approach for evaluating distance to objects, J. Intell. Robot. Syst. 80 (1) (2015) 7–22.

[53] J.A. Marvel, R. Norcross, Implementing speed and separation monitoring in collaborative robot workcells, Robot. Comput.-Integr. Manuf. 44 (2017) 144–155.

[54] D. Rey, E. Papadoupoulos, Online automatic tipover prevention for mobile manipulators, in: Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications, Vol. 3, IROS'97, IEEE, 1997, pp. 1273–1278.

[55] Y. Li, Y. Liu, Real-time tip-over prevention and path following control for redundant nonholonomic mobile modular manipulators via fuzzy and neural-fuzzy approaches, 2006.

[56] K. Alipour, A. Hasanpour, P. Daemy, Comparing two online tip-over avoidance algorithms for mobile manipulators, in: 2014 Second RSI/ISM International Conference on Robotics and Mechatronics, ICRoM, IEEE, 2014, pp. 310–315.

[57] V. Villani, F. Pini, F. Leali, C. Secchi, Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications, Mechatronics 55 (2018) 248–266.

[58] Aarhus University, Robotic colleague takes over the production line at aarhus factory, https://ingenioer.au.dk/en/collaboration/projects/robotic-colleague-takes-over-the-production-line-at-aarhus-factory/.