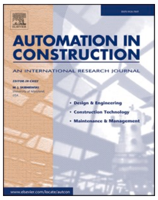
Automation in Construction 141 (2022) 104387

Contents lists available at ScienceDirect

Automation in Construction

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



Robot for automatic waste sorting on construction sites

Xinxing Chen [a,b,1], Huaiyang Huang [a,1], Yuxuan Liu [a,1], Jiqing Li [c,1], Ming Liu [a,d,e,*]

[a] *Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China*
[b] *School of machanical and electrical engineering, Shenzhen Polytechnic, Shenzhen 518055, China*
[c] *Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen University Town, Shenzhen, PR China*
[d] *The Hong Kong University of Science and Technology (Guangzhou), Nansha, Guangzhou 511400, Guangdong, China*
[e] *HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen, China*

ARTICLE INFO

Keywords:
Waste sorting
Construction site
SLAM
Deep learning

ABSTRACT

As a large amount of waste generated from global construction activities, robots that can automatically recycle construction and demolition (C&D) waste have become efficient tools for conserving natural resources, but the complex environment and high diversity of waste on the construction site raise challenges for robot patrolling, object recognition, and grasping. This paper describes a robot for C&D waste recycling, achieving real-time navigation through Simultaneous Localization and Mapping (SLAM). Additionally, a deep learning method and a high-precision 3D object pickup strategy were adopted for the accurate identification and stable grasping of waste items. The recognition accuracy of various kinds of C&D waste was analyzed under different illumination and spatial density conditions. Based on this research, the automation level and the application scenario of the robot prototype would be further improved and broadened.

1. Introduction

Human activities produce a huge amount of waste, which causes severe environmental problems. In 2015, 8.7 billion tons of municipal solid waste were produced worldwide, and the waste quantities continue to grow every year. By the end of the 21 century, it is expected that the amount of global waste will have doubled or even tripled if not managed properly [1]. Construction and demolition (C&D) waste is defined as the surplus or damaged products and materials that arise from construction, renovation, and demolition activities [2]. C&D waste often represents the largest proportion of the total waste generated. For example, in Australia, C&D waste accounts for about 44% of the total amount of annual waste across all industry sectors [3]. Generally, C&D waste is a mixture of inert, non-inert, harmless, and harmful materials, so effective sorting is an essential step in the disposal of C&D waste [4]. However, currently, the C&D waste mixture is usually transported directly to landfills without prior distinction [5,6]. This coarse waste management causes air, water, and soil pollution, and puts tremendous pressure on limited landfill spaces, which has resulted in many serious accidents. For example, in April 2017, at least 30 people died in a landslide at a solid waste landfill site in Sri Lanka, following the death of over 115 people from another landfill landslide in Ethiopia the prior month [7].

As C&D waste could be reused as raw materials, recycling C&D waste is a widely acknowledged way to conserve natural resources. Generally, there are two strategies for C&D waste recycling: off-site recycling and on-site recycling. In most countries, off-site recycling is a more popular strategy in which C&D waste is transported to centralized recycling plants for treatment. However, this strategy has certain drawbacks, such as tremendous transportation costs and high demand for land occupation [8]. In contrast, on-site C&D waste recycling, which sorts waste directly on the construction site, can minimize the cost and pollution issues associated with waste transportation and storage [9]. As such, on-site waste recycling is deemed to be one of the most efficient strategies of waste management [10]. In Japan, the overall recycling rate is higher than 90% as on-site waste sorting is strictly enforced on every construction site. But limited site space, trivial management work, and high labor cost and time consumption collectively hinder the popularization of on-site waste recycling all over the world [11]. Therefore, new automation technology is in high demand to solve these problems of on-site C&D waste sorting.

The past decade has witnessed advances in construction automation.

[*] Corresponding author at: Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China.
E-mail address: eelium@ust.hk (M. Liu).
[1] Indicate equal contribution.

https://doi.org/10.1016/j.autcon.2022.104387
Received 6 September 2021; Received in revised form 18 May 2022; Accepted 24 May 2022
Available online 28 June 2022

As an emerging technology, construction robots play an increasingly important role in the greening of the building sector [12–14]. In particular, off-site waste robotics integrating advanced computer vision and waste handling processes has been developed to sort and recycle bulky C&D waste with high efficiency. Xiao et al. developed a machine to automatically sort C&D waste on a conveyer belt [15]. Kujala et al. designed a truss-type robot for heavy objects grasping on a conveyer belt [16]. In comparison, mature technology for on-site C&D waste recycling has long been absent. Highly autonomous C&D waste recycling on construction sites remains a challenge, mainly due to unstructured construction environments and the wide variety of waste. Recently, robot prototypes for on-site waste recycling have been proposed [17,18]. Computer vision and neural network approaches have been used to identify waste objects [17], while LiDAR-based simultaneous localization and mapping (SLAM) technology is adopted for robot patrolling [18]. However, only two or three types of C&D waste, such as nails and screws, can be recognized. Moreover, an ordinary cam- era is used to determine the 2D position of the target object, which may prevent precise grasping of waste with a complex 3D structure. In addition, the LiDAR- based SLAM algorithm suffers from unsuccessful localization due to the limited information acquired [19]. Therefore, it remains unclear whether automatic waste recycling can be achieved efficiently in a complex environments, such as the construction sites.

Motivated by the urgency for efficient on-site waste recycling, a high-robustness robot prototype has been designed and built, which adopt a novel SLAM method and high-accuracy 3D object picking strategy for automatic C&D waste recycling on construction sites. Specifically, an advanced sensor module is assembled with two RGB-depth (RGB-D) cameras and a 3D LiDAR. The 3D LiDAR and one RGB-D camera at the top of the robot perceive the environmental information for fast and accurate robot localization. In comparison with a one-sensor SLAM algorithm, the integrated LiDAR-camera sensory system takes advantage of both vision-based and range-based methods, allowing real-time localization of the robot, which is especially essential for patrolling around complex environments. Another long-existing challenge for waste collection on construction sites, is the wide variety of waste, with diverse morphology, volume and materials, thus resulting in great difficulty in waste classification and object grasping. To overcome this challenge, a second RGB-D camera is attached to the wrist of a six-degree-of-freedom (6-DOF) robotic arm, serving as an image and depth sensor to scan the ground, recognize waste objects, and determine their 3D positions. Accordingly, an expanded dataset is established by collecting RGB-D images of waste objects captured on construction sites with complex backgrounds. Thanks to this comprehensive dataset, the robot can recognize various waste objects efficiently with the Mask R-CNN (Region-Based Convolutional Neural Networks) algorithm [20]. Based on the depth information captured by the RGB-D camera, the 3D model of the detected waste object can be constructed, and the optimal 3D pickup point can be computed. Then, the robotic manipulator can grasp each waste object and put it into a multi-cell recycling bin on the robot. A powerful and compact control system was built to efficiently drive the robot patrolling, object identification, and grasping functions to achieve automatic waste sorting and recycling on construction sites. The feasibility of the developed robot prototype was verified by both laboratory and field tests. It is demonstrated that this robot has an outstanding capacity to perform automatic C&D waste recognition and collection under complex environments, thus providing a robust and reliable prototype for automatic waste sorting on construction sites.

2. Overview of the proposed robot prototype

2.1. Hardware configuration

The robot prototype for construction waste sorting is shown in Fig. 1. The robot system is composed of five major parts: (1) sensors for environment perception, (2) a robot base, (3) a manipulator, (4) a Next Unit of Computing (NUC), and (5) a waste container.

The sensor module of the robot consists of three environment perception devices, including a 3D Velodyne LiDAR PUCK-16 (VLP-16) and two D435i cameras. LiDAR is a method for determining variable distance by targeting an object with a laser and measuring the time of return for the reflected light. VLP-16 is a small, compact LiDAR optimized for a variety of applications in mapping and robots. It can achieve high accuracy of 3 cm within its working range of about 100 m. The

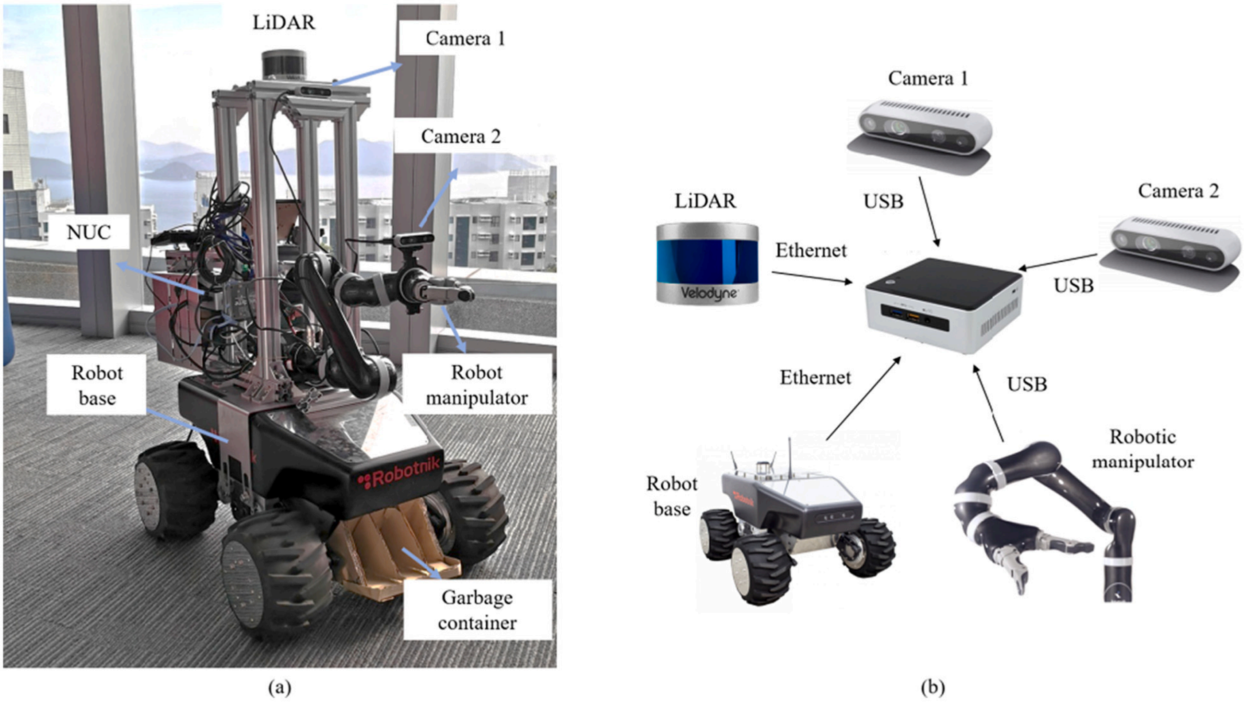


Fig. 1. (a) Prototype of the proposed robot, (b) construction of the prototype.

D435i camera is a cutting-edge stereo depth camera, which can collect RGB images and depth images simultaneously. The depth image presents the distance between the camera and the object with an error of smaller than 2% within a 2 m range. The D435i camera has been widely used for depth sensing thanks to its high resolution, low cost, and small size.

The robot base is Robotnik Summit XL mobile robot with the dimensions 720 × 613 × 392 mm (length × width × height). It is equipped with 4 high-power motor wheels, which can be well adapted to complex outdoor environments with a skid-steering configuration. Its maximum speed is 3 m/s and it can carry up to 20 kg of weight. A motor driver is used to control the rotation, forward, or back translation of the robot.

To manipulate objects of various shapes and sizes, a 6-DOF Kinova MICO2 arm equipped with a 2-finger gripper KG-2 was installed on the robot prototype. Kinova MICO2 is a lightweight humanoid arm, with a total weight of only 4.6 kg. The payload of the manipulator is 1.3 kg.

An Intel NUC minicomputer (Intel(R) Core(TM) i5-6260U CPU @ 1.80GHz, 16GB RAM, 1000 Mb Ethernet) was used to control the robot. All the hardware components were connected to the minicomputer via USB or serial ports without excessive wiring.

A multi-cell garbage container was fixed to the button of the robot base, and used to collect and sort different kinds of C&D waste.

2.2. Software system

A software system was designed and built for automatic patrolling, waste detection, and grasping. As shown in Fig. 2, the software system can be divided into three sub-systems based on their functions: the patrolling system, the waste detection system, and the waste grasping system. Specifically, the patrolling system receives information from Camera 1 and the LiDAR, and then sends the localization information to the robot base, and Camera 2 collects images for the waste detection system. After waste identification, the manipulator is controlled by the waste grasping system to collect the waste object. As shown in Fig. 2, when waste is not detected, the robot base continues the patrolling process, and the manipulator maintains in the holding position. Camera 1 and the LiDAR provide information about the environment for robot localization. During this period, Camera 2 continuously sends images to the waste detection system for object recognition. Once a waste object is detected by the waste detection system, the patrolling process is paused. Then, the waste detection system calculates the 3D coordinates of the pickup point and sends the results to the waste grasping system, which performs the motion planning for waste grasping.

The software system was developed on the robotic operating system (ROS). As shown in Fig. 3, hardware and algorithms are presented as nodes in the graph. They retrieve or pass messages to each other to control the robot's behavior. In brief, visual-LiDAR-localization is a node that localizes the robot on a pre-built map based on the images and point cloud information received from Camera 1 and LiDAR. Then, the output local occupancy map is transferred to the move base node for robot navigation. The images captured by Camera 2 are processed by the object-detection node. When a waste object is detected, its 3D coordinates are calculated and transferred to the Judge node. A message is then broadcast to the base to stop patrolling, followed by a command to make the manipulator grasp the waste object.

3. Key methodology

3.1. SLAM

3.1.1. Overview

To improve the precision of robot localization, a novel 3D localization method was developed by combining both visual and LiDAR information as illustrated in Fig. 4. Two advanced SLAM systems, ORB-SLAM2 [21] and LOAM [22], are employed to build the visual map and the LiDAR map, respectively. ORB-SLAM2 is an indirect visual SLAM method for building a lightweight visual map and localizing the camera. ORB-SLAM2 extracts local keypoints, and computes their descriptors, which are matched against the visual landmark map. The camera pose is tracked with the matched correspondences. In the backend, ORB-SLAM2 jointly optimizes the landmark positions and the camera poses. LOAM is used to build a dense point cloud map. Planar and edge points are extracted from every LiDAR scan and are then aligned against the dense local map to estimate the frame pose. After registration, the current scan is inserted into the local map, which incrementally reconstructs the entire environment.

Then the extrinsic parameters were calibrated, including the relative rotation and translation between the LiDAR and the camera, to generate

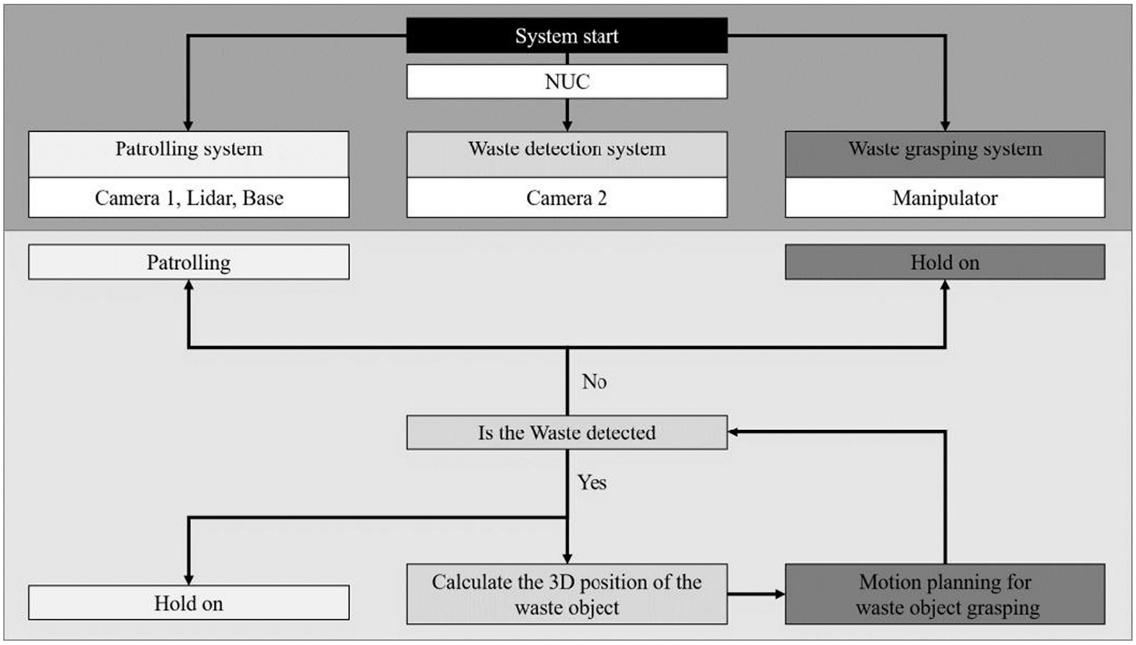


Fig. 2. Architecture of the ROS system.

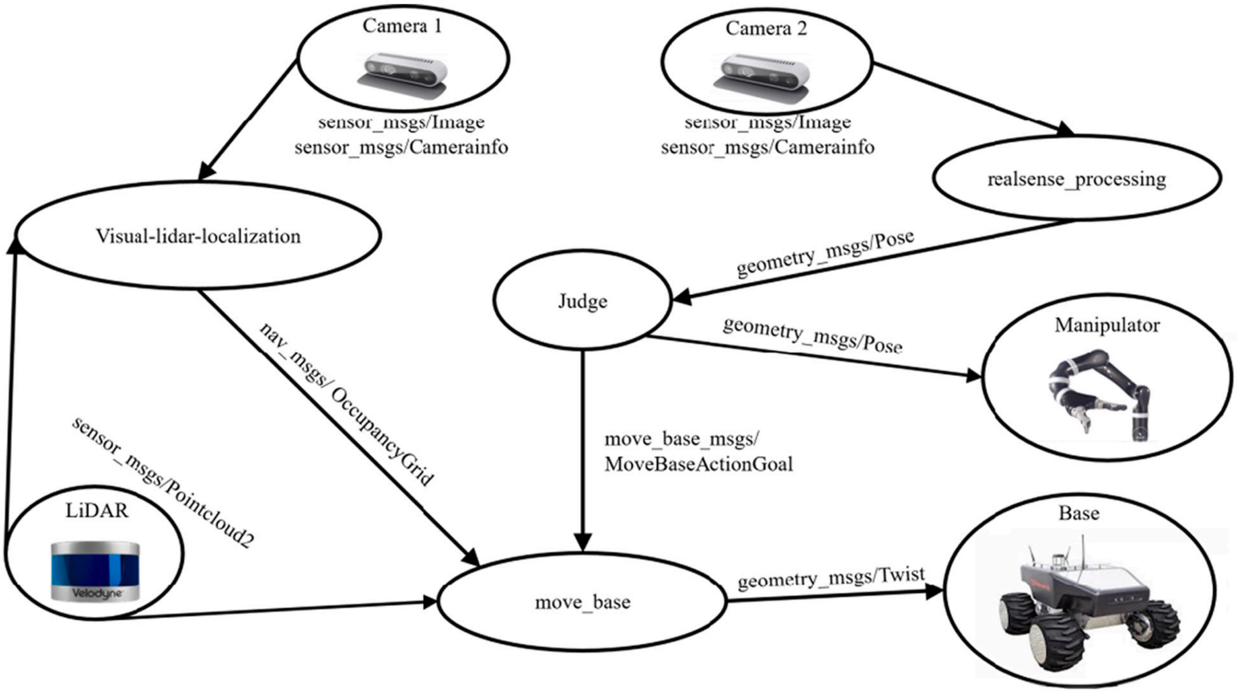


Fig. 3. Structure of the ROS system.

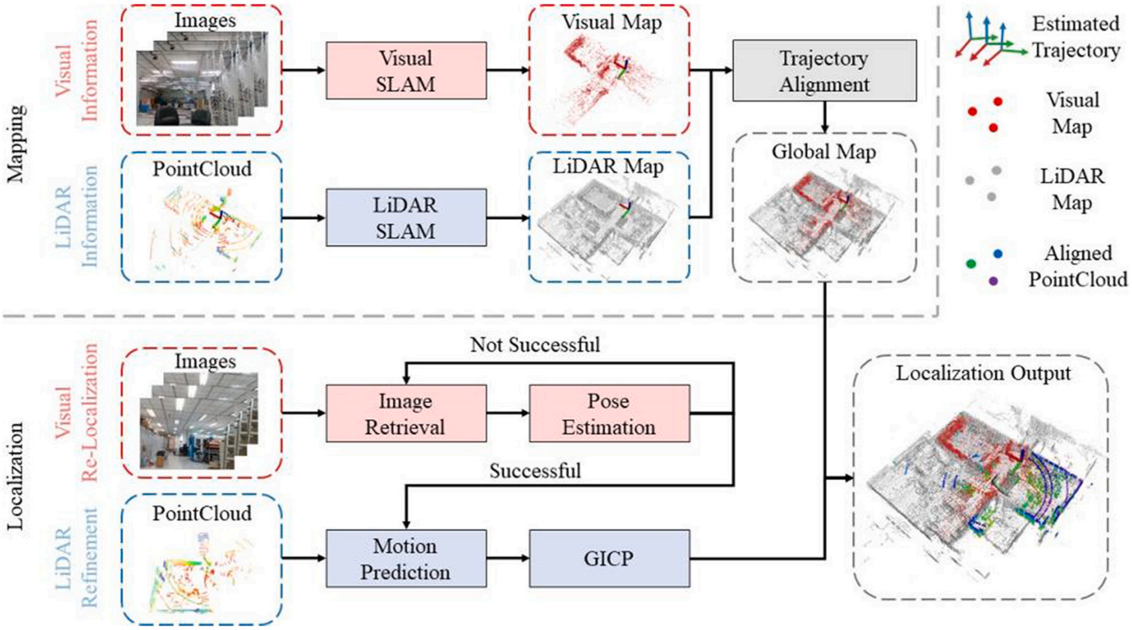


Fig. 4. System flowchart of the localization module.

a consistent map by merging the visual map and the LiDAR map. For robot localization, firstly an initial guess of the robot location is computed from the visual information captured by Camera 1. Sequential images are acquired until the current visual frame is successfully localized on the pre-built map. Then, the visual-based pose is used as the initial guess for LiDAR localization, as shown in Fig. 4. Finally, the current LiDAR scan is registered to the dense map to output a 6-DoF pose estimation for navigation use.

3.1.2. Mapping

In the mapping stage, an integrated map containing both visual and LiDAR information is built by ORB-SLAM2 and LOAM, two state-of-art SLAM methods in visual and LiDAR localization fields, respectively. For map reconstruction, firstly, the visual images and LiDAR information were captured simultaneously in the target scenario. Then, the trajectories of both sensors are computed to generate the visual map and the LiDAR map. The visual map is represented by a bipartite graph of keyframes and visual landmarks. The keyframe generally stores information about the estimated pose and global descriptor for re-localization, while the visual landmark mainly stores its rep- resentative descriptor and position $\boldsymbol{P}_i$. The visual observations are represented by the edges between the landmark and the keyframe. On the other hand, the LiDAR map is simply represented by a point cloud.

The next step is to align the maps of the two modalities that were built separately. The maps can be aligned based on the extrinsic parameters between the camera and LiDAR, i.e., the rigid body translation (RBT) from the camera to the LiDAR $\boldsymbol{T}_{LC}$, from the camera to the world $\boldsymbol{T}_{WC}$, and from the LiDAR to the world $\boldsymbol{T}_{WL}$. Here, an automatic method is employed, which does not need a calibration process in advance. First, the trajectories of two SLAM systems is synchronized with the closest time stamps, which gives two sets of poses, denoted as $P_C = \{\boldsymbol{T}_{WC_k}\}_{k=1\dots n}$and $P_L = \{\boldsymbol{T}_{WL_k}\}_{k=1\dots n}$. Then, the Umeyama method is used to align these two trajectories, which minimizes the following objective function:

$$\widehat{\boldsymbol{T}}_{LC} = arg^{min}_{\boldsymbol{T}_{LC}} \frac{1}{n} \sum_{k=1}^{n} \left\| \mathbf{t}_{WL_k} - \left(\mathbf{R}_{LC} \mathbf{t}_{WC_k} + \mathbf{t}_{LC} \right) \right\|_2^2 \tag{1}$$

where $\mathbf{R}_{LC}$ and $\mathbf{t}_{LC}$ is the rotational and translational component of $\mathbf{T}_{LC}$, respectively. With the extrinsic parameters, the visual map is then transformed under the global frame to ensure consistency between the two maps.

3.1.3. Localization

After map construction, localization procedures are performed hierarchically, in which the initial frame is re-localized coarsely by the visual data and then is fine-tuned by the LiDAR data. This strategy can take advantage of both modalities. The visual localization system can recognize the revisited places and provide the initial guess, while the LiDAR localization systems can register the input scan with high accuracy.

For the visual localization part, the system first extracts Oriented FAST and Rotated BRIEF (ORB), which is a feature descriptor invariant to rotation and scale, from the input image. Next, these features are assigned to the vocabulary tree and are transformed into a global descriptor using the Bag-of-Words (BoW) model [23]. BoW divides the descriptor space into several clusters, and each extracted descriptor is assigned to the closed cluster. The number histogram of descriptor assignment yields a global descriptor of the input image. With the BoW model, the relevant frames are retrieved by the similarity of the global descriptors against the keyframes. Next, for each candidate, the camera pose can be estimated by minimizing the reprojection error between the matched 3D landmarks and the 2D key points, also regarded as the Perspective-n-Point (PnP) problem [24]. According to the guidance of ORB-SLAM2, this problem was solved in a random sample consensus (RANSAC) scheme [25]. The method randomly samples minimal correspondence pairs to estimate the camera pose and count the number of inliers. After a fixed number of iterations, the estimation with maximum inlier correspondences is considered to be the best solution. If the solution is valid, the camera poses can be refined accordingly. Visual localization is performed continuously until the visual frame is successfully matched in the map, and then the pose estimation will be sent to the LiDAR for further localization.

With the initial guess, the LiDAR localization part registers sequential point cloud scans against the prior LiDAR map. For the first frame, the visual localization result was used as the initial guess, given by $\overline{\boldsymbol{T}}_{WL_O} = \boldsymbol{T}_{WC_o} \boldsymbol{T}_{LC}^{-1}$. Forthe subsequent $\boldsymbol{k}$th frame, a constant velocity model is assumed and the pose is given by

$$\overline{\boldsymbol{T}}_{WL_k} = \boldsymbol{T}_{WL_{k-1}} \boldsymbol{T}_{WL_{k-2}}^{-1} \boldsymbol{T}_{WL_{k-1}} \tag{2}$$

Then, the Generalized Iterative Closest Point (GICP) [26] is used as the registration method for pose fine-tuning. Compared to the traditional ICP method, GICP assigns each point with a Covariance matrix. Similar to ICP variants, GICP iteratively associates points between the source and the target point cloud and estimates the relative pose weighed by the covariance of each point. The original GICP method uses a KD-Tree-based association strategy, where a binary tree is built with each node dividing one of the three dimensions into two half-spaces. This KD-Tree-based [27] association requires high computational resources for correspondence estimation. To alleviate the computational cost and perform scan-matching in real-time, a voxel-based implementation is adopted [28], where the information of components is stored in voxels with unique indices hashed from the voxel location. In addition to a voxel-based GICP implementation, two strategies are used to make the LiDAR localization module real-time. Firstly, the input point cloud is downsampled using a voxel filter with a specific resolution. Secondly, voxelization and covariance estimation are performed in advance of the scan matching.

3.2. Picking strategy

A robust picking strategy for C&D waste objects of varied shapes is established. The picking strategy is divided into three parts: object detection, 3D coordinate estimation, and coordinate transformation. Firstly, the Mask R- CNN method [20] is applied to detect the object's boundary. Mask R-CNN is a deep neural network architecture that aims to solve instance segmentation problems in computer vision. It can identify an object's class, draw the bounding box, and delineates an object's boundary. Then, the RGB-D camera is used to get both the RGB images and the corresponding aligned depth images for 3D coordinate estimation of the center of gravity. Finally, the estimated 3D coordinate is transformed from the camera coordinate to the base of the robot manipulator for grasping pose estimation.

3.2.1. Object recognition

To ensure precise object grasping, firstly the boundary of the waste object should be delineated at pixel-level accuracy. In this study, instance segmentation, a deep learning technology, was adopted to detect the contours of all objects in the image. Among the state-of-the-art algorithms for instance segmentation, Mask R-CNN surpasses other algorithms, such as MNC[29] and FCIS+++[30], in terms of accuracy in various categories [20]. The basic structure of the Mask R-CNN is shown in Fig. 5. Firstly, the input image is sent to the region proposal network (RPN), which consists of a ResNet backbone network, feature pyramid network, and class-agnostic detection head. The RPN outputs multiple candidate bounding boxes. Then, features in the proposed regions are aligned and fed to additional prediction branches. A classification branch is applied to predict the class of the object and output the class label. Meanwhile, a bounding-box regression was performed for the candidate box to get the offset for bounding boxes refinement. Within the box boundary, a segmentation mask is then predicted in a pixel-to-pixel manner using a fully convolutional network (CONV). A great challenge for precise detection of C&D waste objects is that the background and light conditions on construction sites are largely variable and distinct. Therefore, to improve the applicability of the proposed method in field tests, it is necessary to create a dataset of images collected under real construction site circumstances.

In detail, the newly collected construction waste dataset has 756 RGB images with the same resolution of 640 × 480 pixels. The dataset contains seven classes of waste objects, including cotton gloves, wood blocks, small ferrous, plastic pipe, bamboo, corrugated paper, and rebar. All the images were taken on a construction site covering large variations in scenes and illumination conditions. Each waste object in the dataset was annotated by classical pixel-wise segmentation with a class label, as shown in Fig. 6. The 756 images were split into 454 identities for training, 151 identities for validation, and the remaining 151 identities for testing. The dataset was trained using the mmdetection2 platform. Table 1 shows the mAP (mean averaged over IoU thresholds) at the Intersection-over-Union (IoU) metrics from 0.5 to 0.95. As can be seen, the mAP for box detection and segmentation approaches 0.66 and 0.68, respectively. Meanwhile, the Average Recall (ARs) for box detection and segmentation is 0.70 and 0.72, respectively. The mAP (IoU = 0.5:0.95) for different kinds of objects is shown in Table 2.

The representative outputs of Mask R-CNN are demonstrated in

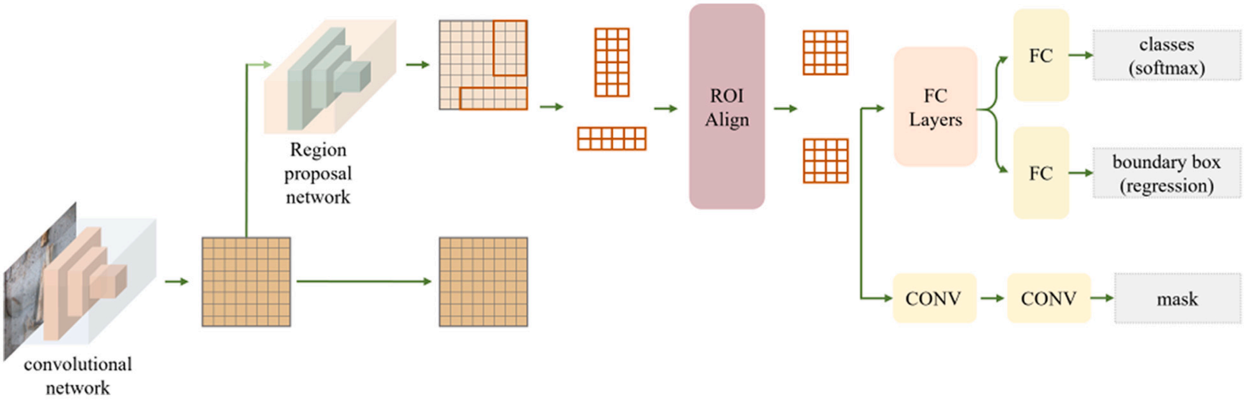
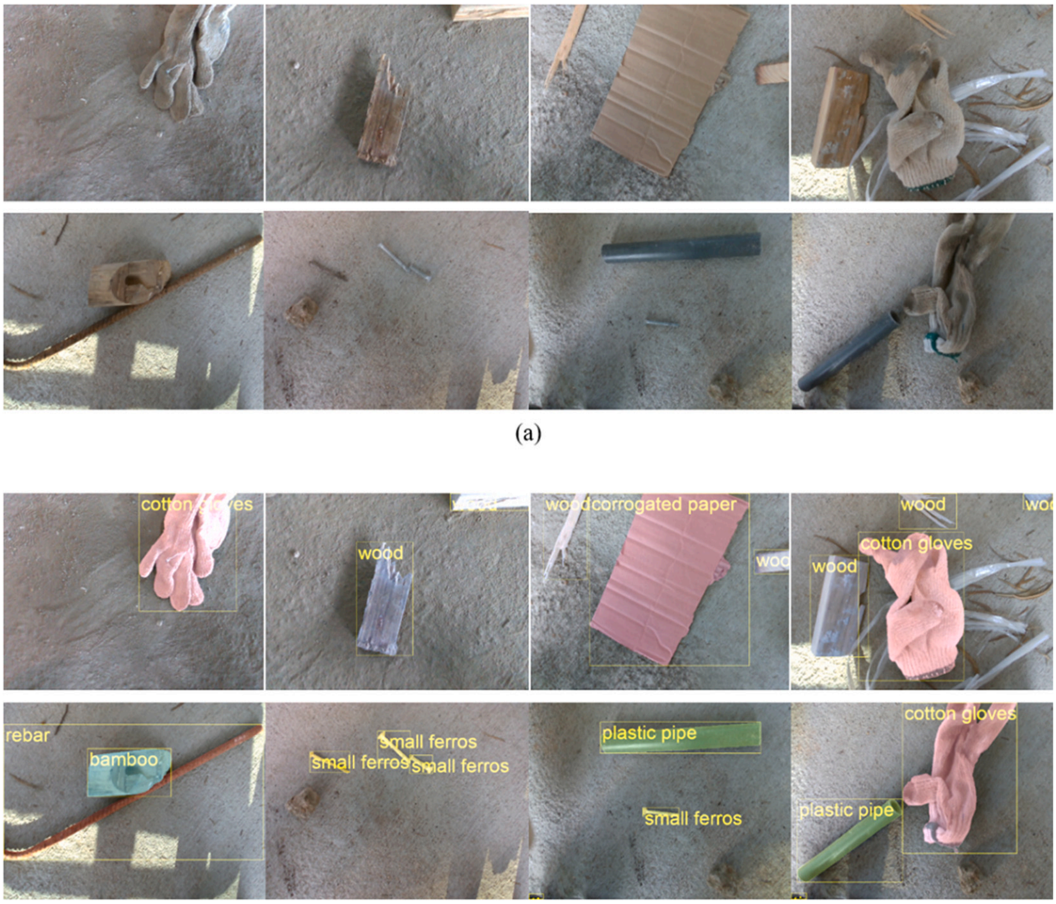


Fig. 5. MaskR-CNN framework.



(b)

Fig. 6. Data examples, (a) RGB images, (b) corresponding label images.

Table 1
mAP for box and segmentation.

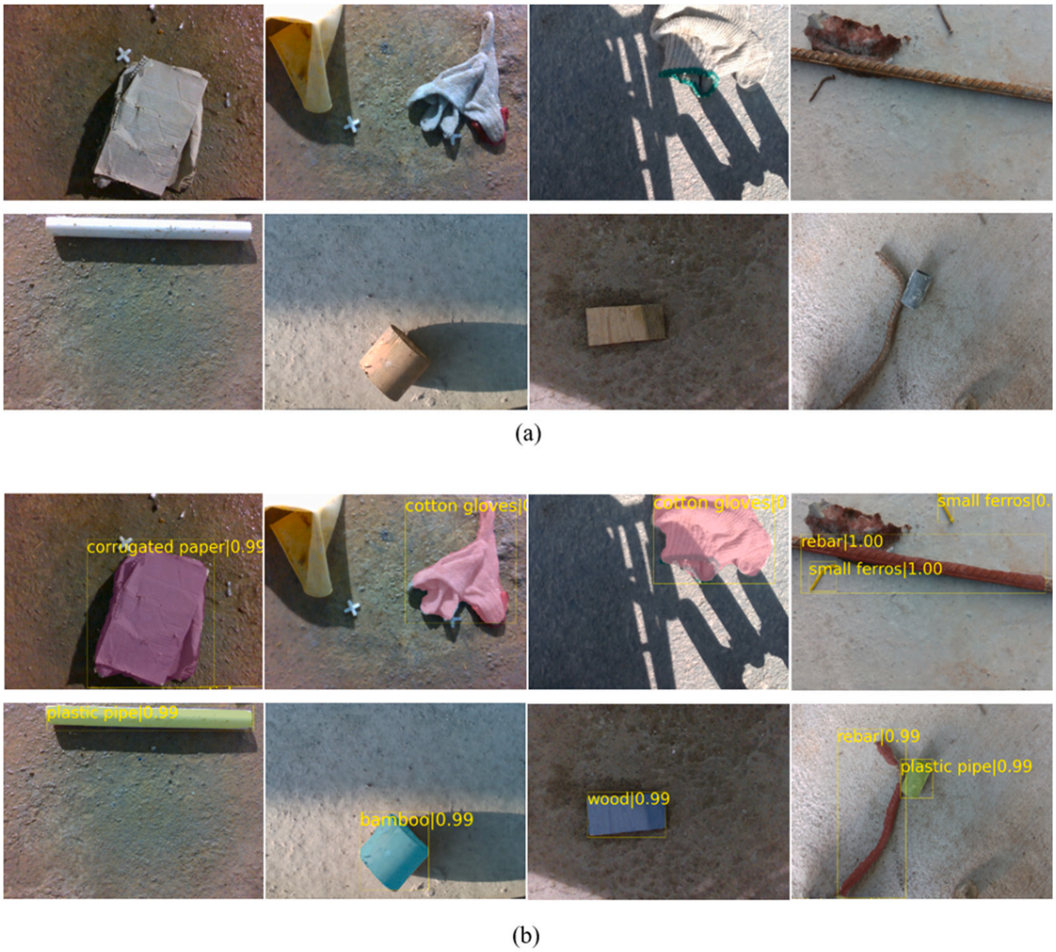
IoU			
mAP	0.5:0.95	0.5	0.75
segm	0.683	0.867	0.772
box	0.657	0.865	0.788

Fig. 7. As can be seen, Mask R-CNN can precisely identify various kinds of objects with diverse shapes and colors, even under complex light conditions and with different backgrounds, suggesting the high robustness of the Mask R-CNN model.

Next, an in-depth evaluation was conducted to evaluate the recognition accuracy in different site environments. Specifically, the waste images were classified into three classes according to different spatial densities and light contrasts, as shown in Fig. 8. Quantitative analysis of the segmentation results Table 3 in shows that Mask R-CNN can recognize waste targets under different light conditions, with the mAP values

Table 2
mAP for different classes of objects.

mAP	Different classes of objects						
	Cotton gloves	Wood block	Small ferros	Plastic pipe	Bamboo Bamboo	Corrugated Paper	Steel Bar
segm	0.734	0.711	0.405	0.593	0.805	0.952	0.771
box	0.666	0.737	0.549	0.584	0.721	0.768	0.782



(a)

(b)

Fig. 7. Instance segmentation results, (a) RGB images in the test datasets, (b) corresponding predicted images.

of high light-dark contrast images slightly lower than those of medium and low light-dark contrast ones, possibly due to the highly heterogeneous brightness caused by light and shadows. The recognition accuracy increases with the decreasing spatial density of the waste, as shown in Table 3. These results indicate that the Mask R-CNN algorithm is highly robust to different site environments.

3.2.2. 3D coordinate estimation

In order to pick up waste objects stably, the center of gravity should be calculated with high accuracy, which is presented as the pickup point ***P***in Fig. 9. To locate the point ***P***, the 3D model of the waste object is firstly built prior to the calculation of the 3D coordinate of the center of gravity (X_C, Y_C, Z_C). Next, the point ***P*** is transformed from the camera coordinate to the robot manipulator coordinate (X_R, Y_R, Z_R) using $\boldsymbol{R}_{RtoC}$ and $\boldsymbol{T}_{RtoC}$, which are the external parameters connecting the two coordinate systems.

3D model reconstruction. Prior to 3D coordinate estimation, 3D model reconstruction should be conducted based on the RGB and depth images captured by the RGB-D camera, as shown in Fig. 10(a) and (b). Firstly, a median filter is used to make the depth image more smooth (Fig. 10(c)). Then the depth image is transformed to the 3D point cloud by using the calibrated internal parameters of the camera [31], as shown in Fig. 10 (d). Next, with the help of instance segmentation results obtained from Mask R-CNN (Fig. 10(e)), the point cloud of the detected waste objects can be separated from that of the background, which is marked in yellow and blue respectively in Fig. 10(f), respectively.

Computation for center of gravity. The center of gravity ***P***, which is presented as $\overline{X_C}$, $\overline{Y_C}$, and $\overline{Z_C}$, is calculated by the following equations:

$$\overline{X_C} = \frac{\iiint_D X_C d\sigma}{\iiint_D d\sigma} \tag{3}$$

$$\overline{Y_C} = \frac{\iiint_D Y_C d\sigma}{\iiint_D d\sigma} \tag{4}$$

$$\overline{Y_C} = \frac{\iiint_D Y_C d\sigma}{\iiint_D d\sigma} \tag{5}$$

in which D is the total volume of the detected waste object, and dσ is a unit volume in D. After calculation of the center of gravity $\boldsymbol{P}(\overline{X_C}, \overline{Y_C}, \overline{Z_C})$,

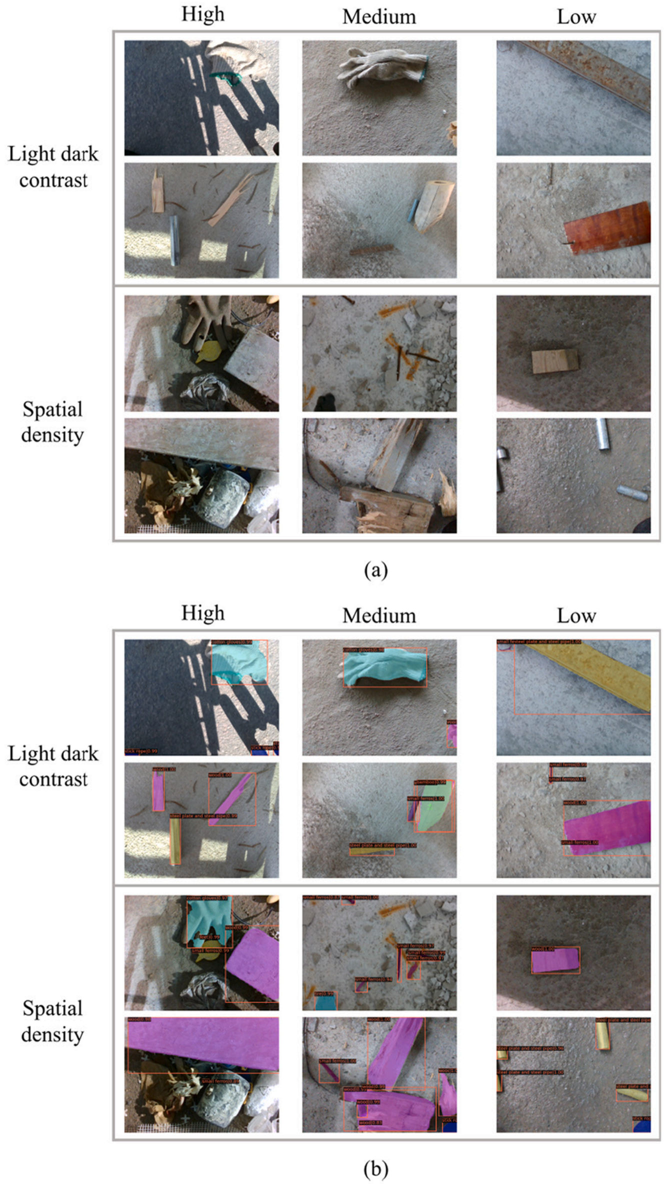


Fig. 8. Instance segmentation results of C&D waste under different site environments. (a) RGB images with different light-dark contrast and spatial density, (b) corresponding predicted images.

Table 3
mAP for waste recognition under different site conditions.

mAP	Spatial density			Light-dark contrast		
	High	Medium	Low	High	Medium	Low
segm	0.571	0.653	0.707	0.565	0.664	0.672
box	0.618	0.676	0.719	0.611	0.701	0.704

the pickup direction is determined as the shortest line segment that passes through the point ***P***.

3.2.3. Coordinate transformation

After obtaining the pickup point ***P*** and the pickup direction in the camera coordinate system, they are transferred to the coordinate system of the robot manipulator with the following equation:

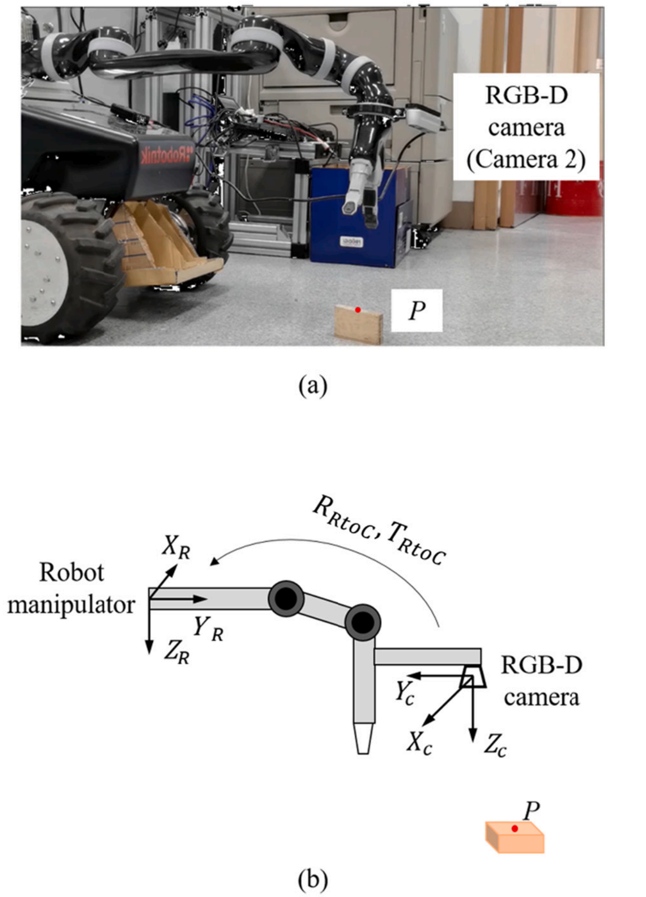


Fig. 9. 3D coordinate estimation: (a) photography of the robot manipulator for waste grasping, (b) schematics of the robot manipulator and 3D coordinate systems.

$$\begin{bmatrix}X_R\\Y_R\\Z_R\end{bmatrix} = R_{RtoC}\times\begin{bmatrix}X_C\\Y_C\\Z_C\end{bmatrix}+T_{RtoC} \tag{6}$$

where R_{RtoC} and T_{RtoC} is the rotation matrix and translation matrix, respectively, from the robot manipulator coordinate system to the camera coordinate system. They are obtained using the eye-in-hand calibration method [32,33].

4. Experiments

4.1. Laboratory test

To evaluate the performance of the robot prototype, a laboratory test was conducted (Fig. 11(a)). Firstly, the camera and LiDAR were used to scan the laboratory and construct a visual map (Fig. 11(b)) and a LiDAR map (Fig. 11(c)). Then a merged global map was constructed by using the extrinsic parameters between the camera and the LiDAR (Fig. 11(d)). Next, the 3D global map was transformed to the 2D map, and the extended BSA method was then applied [34] to do full coverage path planning, as shown in Fig. 12(a). During automatic patrolling (Fig. 12(b)), Camera 1 and the LiDAR continuously perceived the environment and sent the information to the SLAM system for robot pose estimation. As shown in Fig. 12(c), the visual frame was successfully re-localized as the feature points of the input image matched with the visual landmarks on the keyframe. Then the LiDAR localization was conducted for fine-tuning of the robot's pose. It was found that the proposed SLAM method can register the sequential point cloud with the pre-built dense

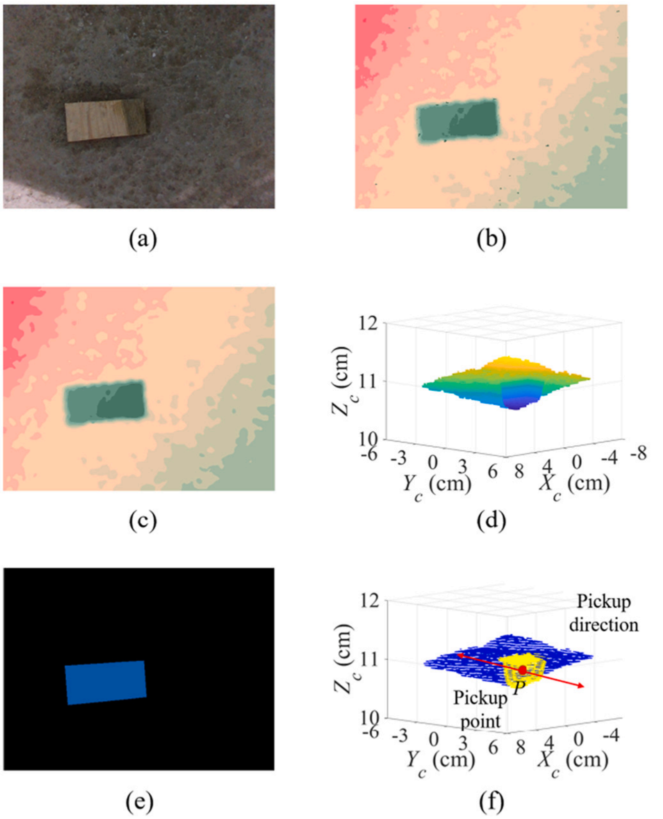


Fig. 10. 3D reconstruction and pickup point calculation: (a) RGB image and (b) corresponding depth image of the detected waste object, (c) smoothed depth image, (d) 3D model of the waste object, (e) the segmented waste object, (f) pickup direction and point of the waste object. X_C, Y_C, and Z_C are the three axes of the camera coordinate).

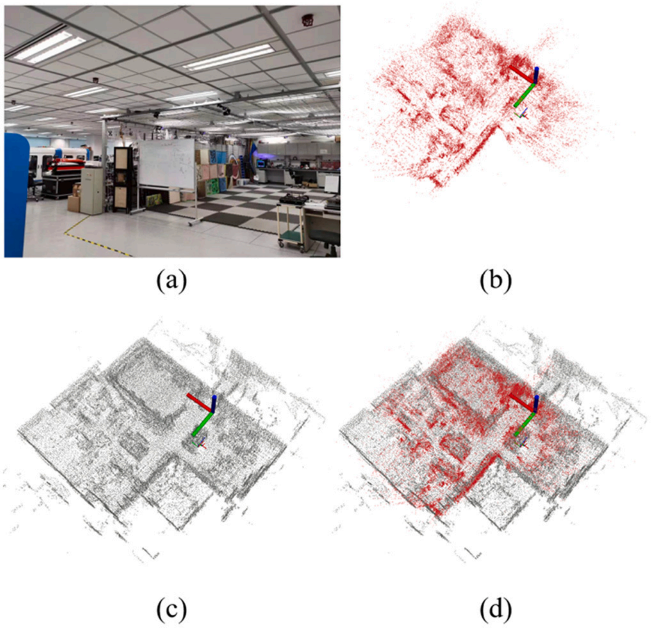


Fig. 11. Map construction for the laboratory experiment: (a) environment of the laboratory; (b) visual map; (c) LiDAR map; (d) global map merged by visual map and LiDAR map.

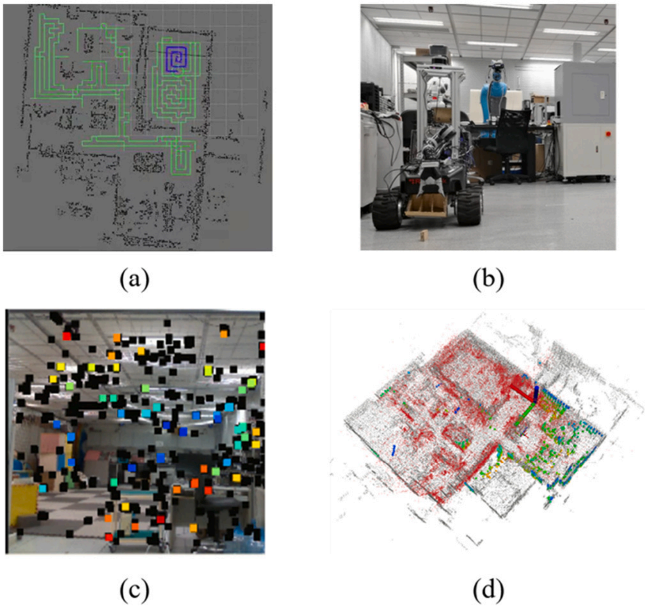


Fig. 12. Patrolling and localization for the laboratory experiment: (a) planning path for the laboratory test; (b) the patrolling robot; (c) visual frame with matched features; (d) localized LiDAR map.

map with high accuracy in real-time. As shown in Fig. 12(d), the green point clouds received in real-time were well-aligned with the point clouds in the pre-built LiDAR map.

At the same time, during the patrol of the robot prototype, Camera 2 sent sequential images to the waste recognition system for automatic detection of waste objects on the path. Once a waste object was detected, its 3D coordinate of the center of gravity was calculated, and the robot manipulator performed motion planning for object grasping. As shown in Fig. 13(a), the robot manipulator was in the holding pose. In Fig. 13 (b), the manipulator grasped the waste object successfully. Then the waste objects were sent to the dustbin automatically, as shown in Fig. 13 (c). After the process of waste sorting, the robot arm returned to its original pose for waste objects detection (Fig. 13(d)), and the robot continued its patrolling process.

4.2. Field test

A field test was conducted to evaluate the outdoor performance of

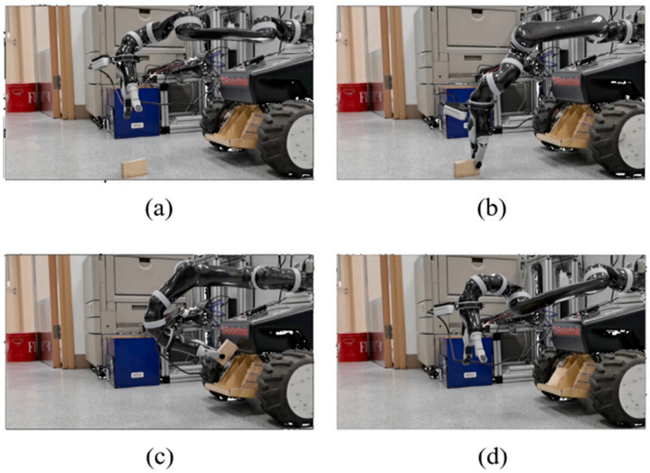


Fig. 13. Automatic picking process for the laboratory experiment: (a) original holding pose; (b) grasping pose; (c) dustbin pose; (d) holding pose.

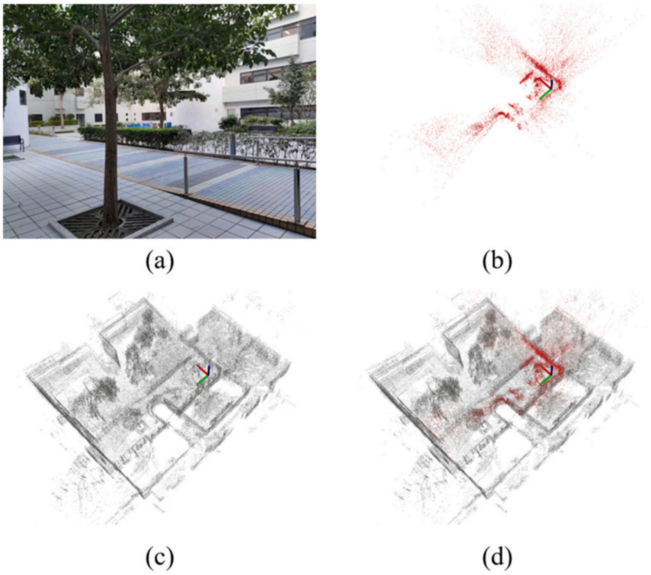


Fig. 14. Map construction for field test: (a) environment of the field; (b) visual map; (c) LiDAR map; (d) global map merged by visual map and LiDAR map.

the pro- posed robot prototype. When the robot entered the outdoor environment (Fig. 14(a)), it used the camera and the LiDAR to scan the site and build the visual map and the LiDAR map, as shown in Fig. 14(b) and (c), respectively. After the estimation of the extrinsic parameters between the cam- era and the LiDAR, a 3D global map was obtained, as shown in Fig. 14(d). Based on the 3D global map, the 2D global map was computed after compression. Then the extended BSA method was applied to perform full coverage path planning. The planed path is shown in Fig. 15(a). Based on the pre-built map and planned path, the robot prototype started to patrol around the out- door environment, as shown in Fig. 15(b). The 3D localization module also started to work. Firstly, the initial pose was calculated based on the visual features, as shown in Fig. 15(c). Then the following pose was successfully traced by the LiDAR localization part, which is shown in Fig. 15(d).

During the patrol of the robot prototype, the robot manipulator was in the holding position, and Camera 2 scans the ground for automatic waste objects detection, as shown in Fig. 16(a). Once a waste object was detected, the patrolling process was paused, and the waste grasping system started its motion planning. Firstly, the pickup point and direction were calculated. Then the robot prototype grasped the waste object and sent it to the dustbin, as shown in Fig. 16(b) and (c), respectively. After sorting the detected waste object, the robot manipulator returned to the holding position (Fig. 16(d)).

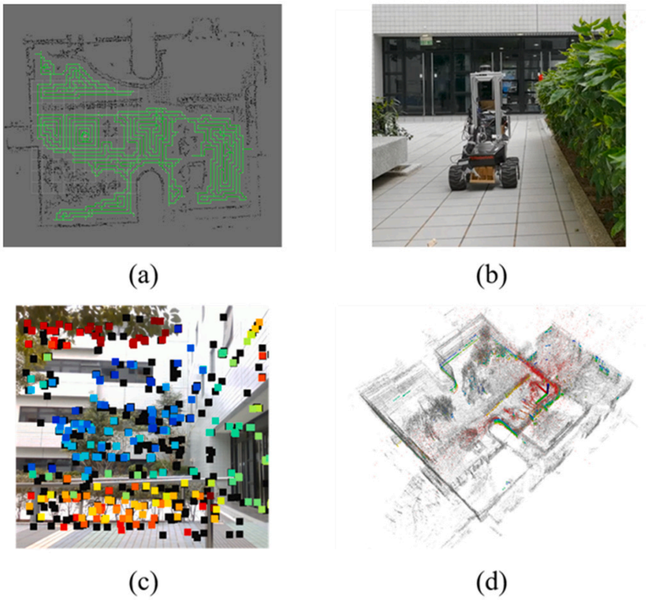


Fig. 15. Patrolling and localization for the field test: (a) planning path; (b) the patrolling robot; (c) visual frame with matched features; (d) localized LiDAR map.

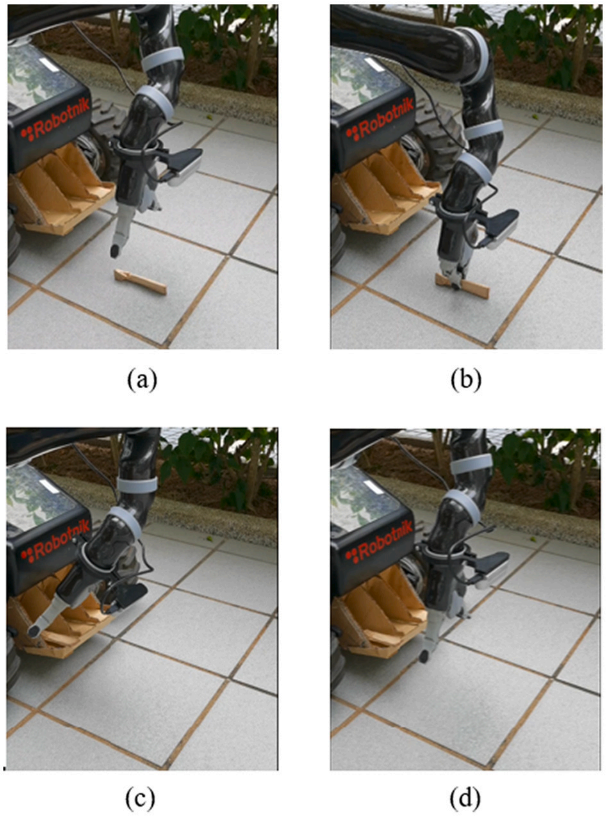


Fig. 16. Patrolling and localization for the field test, (a) planning path for the field test, (b) the patrolling robot, (c) visual frame with matched features, (d) localized LiDAR map.

5. Conclusions

In this study, a robot prototype was developed for automatic waste recycling on construction sites. The main contribution of this study resides in the fol- lowing aspects: (1) build a compact robot prototype equipped with a high-load robot base, high-power motor wheels, and high-performance sensory modules for waste collection under complex outdoor circumstances; (2) develop an automatic patrolling system based on a two-sensor (RGB-D camera and 3D LiDAR) SLAM strategy for real-time 3D localization and navigation with high efficiency and accuracy; (3) train a Mask R-CNN-based recognition model using an expanded image dataset of different types of C&D waste objects captured on real construction sites; (4) establish a high-precision 3D grasping strategy by calculating the 3D coordinates of waste objects based on depth images.

This robot integrates multiple functions, including map reconstruction, navigation, re-localization, waste objects detection, and sorting. Especially, the newly developed SLAM method integrates the visual and LiDAR information for global localization, which aims to estimate the pose of the robot in a pre-built map without any prior knowledge of its initial pose. Global localization can be divided into two stages: initial localization, and pose tracking. LiDAR can track the robot's position with high accuracy in a prior map [35]. However, in the stage of initial localization, it is difficult for the LiDAR to recognize a place that has been seen before with only a limited number of LiDAR scans. In contrast, vision-based methods usually show better performance in this specific stage as rich visual information can be utilized for feature matching. In the second stage for pose tracking, however, vision-based methods have their problems; their computational cost is too high to search for the exact feature correspondences as there are thousands of 3D points and associated feature descriptors in the whole space. Therefore, for continuous pose tracking in the second stage, high-accuracy LiDAR-based methods are preferred. To take full advantage of the two kinds of information, they were integrated for robot localization. Firstly, both the camera and the LiDAR were used to construct a visual map and a LiDAR map, respectively. These two maps are merged to generate a global map after calibration. Next, in the following global localization stage, the visual information is utilized to recover the initial pose of the robot, and the LiDAR localization part registers sequential point cloud scans against the pre-built global map.

C&D waste objects are of varying 3D shapes, making them difficult to grasp by robot manipulator. Traditional 2D image-based detection methods are applicable to objects of similar dimensions, and usually require a fixed distance between the manipulator and the object. However, the environments of construction sites are extremely complex in terms of both waste types and ground conditions. As such, the proposed 3D detection and picking strategy based on depth images are of great significance to increase the robustness and precision of C&D waste sorting and collection. Successful laboratory and field tests demonstrate that the proposed robot prototype can serve as a powerful tool for automatic collection of C&D waste, which would greatly facilitate efficient waste recycling and preservation of natural resources.

The accuracy of target recognition decreases with the increased light-dark contrast and spatial density of objects. Therefore, in case of poor lighting conditions, it is advisable to assemble additional light sources on the robot to create a relatively homogeneous lighting background. As regards the situation of cluttered objects, a sequential grasping scheme can be adopted based on their distance from the robot.

In future work, the robot functions, especially the automatic patrolling and waste grasping processes, will be fully evaluated on actual construction sites under different conditions. Additionally, the application scenarios of this robot will not be limited to construction sites but will be further extended to public places, such as markets, subway stations, and airports. Accordingly, the types of objects for recycling are no longer limited to the C&D waste but will cover more domestic garbage, such as plastic bags, bottles, and so on. A variety of waste images in different architectural scenes should be collected to establish a large dataset. Coupled with optimized Mask R-CNN algorithms, it is applicable for the recognition of a wide spectrum of waste objects under a complex environment. On the other hand, the wide applications of construction robots in public places also raise higher requirements for the grasping system. An improved motion planning scheme should be introduced to avoid any collision with surrounding obstacles or people while grasping the target object. Highly repetitive scenes would degrade the performance of place recognition and localization during robot patrolling. Thus, other sensors, such as GPS, can be integrated into the sensory module to promote robot navigation in both outdoor and indoor environments.

Author contributions

Xinxing Chen: Conceptualization and organization of the entire project; hardware assembly of the robot; design of the 3D picking strategy; manuscript writing and editing; Huaiyang Huang: development of the SLAM method; Yux- uan Liu: software assembly of the robot; Jiqing Li: managing the robot manip- ulator; Ming Liu: conceptulization and supervision of the project, and funding acquisition.

Supporting information

This work was supported by Collaborative Research Fund by Research Grants Council Hong Kong, under Project No. C4063-18G, Department of Science and Technology of Guangdong Province Fund, under Project No. GDST20EG54 and Zhongshan Municipal Science and Technology Bureau Fund, under project ZSST21EG06, Pudong New Area Science and Technology Development Fund, No.PKX2019-R18, awarded to Prof. Ming Liu.

Declaration of Competing Interest

The authors declare no potential conflicts of interests.

References

[1] D.C. Wilson, L. Rodic, P. Modak, R. Soos, A. Carpintero, K. Velis, M. Iyer, O. Simonett, Global Waste Management Outlook, 2015 (ISBN: 978–92–807-3479-9).

[2] V.G. Ram, Satyanarayana N. Kalidindi, Estimation of construction and demolition waste using waste generation rates in Chennai, India, Waste Manag. Res. 35 (6) (2017) 610–617, https://doi.org/10.1177/0734242X17693297.

[3] J. Park, R. Tucker, Overcoming barriers to the reuse of construction waste material in Australia: a review of the literature, Int. J. Constr. Manag. 17 (3) (2017) 1–10, https://doi.org/10.1080/15623599.2016.1192248.

[4] X. Chen, W. Lu, Identifying factors influencing demolition waste generation in Hong Kong, J. Clean. Prod. 141 (2017) 799–811, https://doi.org/10.1016/j.jclepro.2016.09.164.

[5] H. Duan, J. Li, Construction and demolition waste management: China's lessons, Waste Manag. Res. 34 (5) (2016) 397–398, https://doi.org/10.1177/0734242X16647603.

[6] M.D. Bovea, J.C. Powell, Developments in life cycle assessment applied to evaluate the environmental performance of construction and demolition wastes, Waste Manag. 50 (4) (2016) 151–172, https://doi.org/10.1016/j.wasman.2016.01.036.

[7] H. Duan, J. Li, G. Liu, Growing threat of urban waste dumps, Lect. Notes Comput. Sci 546 (7660) (2017), https://doi.org/10.1038/546599b pp.599–599.

[8] Z. Bao, W. Lu, B. Chi, H. Yuan, J. Hao, Procurement innovation for a circular economy of construction and demolition waste: lessons learned from Suzhou, China, Waste Manag. 99 (2019) 12–21, https://doi.org/10.1016/j.wasman.2019.08.031.

[9] Z.K. Bao, W.M.W. Lee, W.S. Lu, Implementing on-site construction waste recycling in Hong Kong: barriers and facilitators, Sci. Total Environ. 747 (2020), https://doi.org/10.1016/j.scitotenv.2020.141091. Article Number: 141091.

[10] Z.K. Bao, W.S. Lu, Developing Efficient Circularity for Construction and Demolition Waste Management in Fast Emerging Economies: Lessons Learned from Shenzhen, China 724, 2020, https://doi.org/10.1016/j.scitotenv.2020.138264. Article Number: 138264.

[11] C.S. Poon, T.W. Ann, L.H. Ng, On-site sorting of construction and demolition waste in Hong Kong, Resour. Conserv. Recycl. 32 (2) (2001) 157–172, https://doi.org/10.1016/S0921-3449(01)00052-0.

[12] T. Bock, The future of construction automation: technological disruption and the upcoming ubiquity of robotics, Autom. Constr. 59 (2015) 113–121, https://doi.org/10.1016/j.autcon.2015.07.022.

[13] Q. Chen, B.G. de Soto, B.T. Adey, Construction automation: research areas, industry concerns and suggestions for advancement, Autom. Constr. 94 (2018) 22–38, https://doi.org/10.1016/j.autcon.2018.05.028.

[14] N. Melenbrink, J. Werfel, A. Menges, On-site autonomous construction robots: towards unsupervised building, Autom. Constr. 119 (2020) 1–21, https://doi.org/10.1016/j.autcon.2020.103312. Article Number: 103312.

[15] W. Xiao, J. Yang, H. Fang, J. Zhuang, Y. Ku, X. Zhang, Development of an automatic sorting robot for construction and demolition waste, Clean Techn. Environ. Policy 22 (9) (2020) 1829–1841, https://doi.org/10.1007/s10098-020-01922-y.

[16] J.V. Kujala, T.J. Lukka, H. Holopainen, Classifying and sorting cluttered piles of unknown objects with robots: A learning approach, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2016, pp. 971–978, https://doi.org/10.1109/IROS.2016.7759167.

[17] Z. Wang, H. Li, X. Zhang, Construction waste recycling robot for nails and screws: computer vision technology and neural network approach, Autom. Constr. 97 (2019) 220–228, https://doi.org/10.1016/j.autcon.2018.11.009.

[18] Z. Wang, H. Li, X. Yang, Vision-based robotic system for on-site construction and demolition waste sorting and recycling, J. Build. Eng. 32 (2020), https://doi.org/10.1016/j.jobe.2020.101769. Article Number: 101769.

[19] Y.J. Lee, B.D. Yim, J.B. Song, Mobile robot localization based on effective combination of vision and range sensors, Int. J. Control. Autom. Syst. 7 (2009) 97–104, https://doi.org/10.1007/s12555-009-0112-0.

[20] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969, https://doi.org/10.1109/ICCV.2017.322.

[21] R. Mur-Artal, J.D. Tardós, ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras, IEEE Trans. Robot. 33 (5) (2017) 1255–1262, https://doi.org/10.1109/TRO.2017.2705103.

[22] J. Zhang, S. Singh, LOAM: Lidar odometry and mapping in real-time, Robotics 2 (9) (2014) 1–9, https://doi.org/10.15607/RSS.2014.X.007. July.

[23] J. Dai, K. He, J. Sun, Instance-aware semantic segmentation via multi-task network cascades, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3150–3158, https://doi.org/10.1109/CVPR.2016.343.

[24] Y. Li, H. Qi, J. Dai, X. Ji, Y. Wei, Fully convolutional instance-aware semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2359–2367, https://doi.org/10.1109/CVPR.2017.472.

[25] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395, https://doi.org/10.1145/358669.358692.

[26] A. Segal, D. Haehnel, S. Thrun, Generalized-ICP, Robotics 2 (4) (2009, June) 435–442, https://doi.org/10.15607/RSS.2009.V.021.

[27] J.L. Bentley, Multidimensional binary search trees used for associative searching, Commun. ACM 18 (9) (1975) 509–517, https://doi.org/10.1145/361002.361007.

[28] K. Koide, M. Yokozuka, S. Oishi, A. Banno, Voxelized GICP for fast and accurate 3D point cloud registration, in: IEEE International Conference on Robotics and Automation (ICRA), 2021, pp. 11054–11059, https://doi.org/10.1109/ICRA48506.2021.9560835. May.

[29] J. Dai, K. He, J. Sun, Instance-aware semantic segmentation via multi-task network cascades, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3150–3158, https://doi.org/10.1109/CVPR.2016.343.

[30] Y. Li, H. Qi, J. Dai, X. Ji, Y. Wei, Fully convolutional instance-aware semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2359–2367, https://doi.org/10.1109/CVPR.2017.472.

[31] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. Pattern Anal. Mach. Intell. 22 (11) (2000) 1330–1334, https://doi.org/10.1109/34.888718.

[32] Y.C. Shiu, S. Ahmad, Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX= XB, IEEE Trans. Robot. Autom. 5 (1) (1987) 16–29, https://doi.org/10.1109/70.88014.

[33] R.Y. Tsai, R.K. Lenz, A new technique for fully autonomous and efficient 3D robotics hand/eye calibration, IEEE Trans. Robot. Autom. 5 (3) (1989) 345–358, https://doi.org/10.1109/70.34770.

[34] E. Gonzalez, O. Alvarez, Y. Diaz, C. Parra, C. Bustacara, BSA: A complete coverage algorithm, in: Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005, pp. 2040–2044, https://doi.org/10.1109/ROBOT.2005.1570413. April.

[35] Z. Su, X. Zhou, T. Cheng, H. Zhang, B. Xu, W. Chen, Global localization of a mobile robot using LiDAR and visual features, in: IEEE International Conference on Robotics and Biomimetics (ROBIO), 2017, pp. 2377–2383, https://doi.org/10.1109/ROBIO.2017.8324775. December.