ORIGINAL ARTICLE

Glass Recognition and Map Optimization Method for Mobile Robot Based on Boundary Guidance

Yong Tao[1,2]* , He Gao[2], Yufang Wen[1], Lian Duan[3] and Jiangbo Lan[3]

Abstract

Current research on autonomous mobile robots focuses primarily on perceptual accuracy and autonomous performance. In commercial and domestic constructions, concrete, wood, and glass are typically used. Laser and visual mapping or planning algorithms are highly accurate in mapping wood panels and concrete walls. However, indoor and outdoor glass curtain walls may fail to perceive these transparent materials. In this study, a novel indoor glass recognition and map optimization method based on boundary guidance is proposed. First, the status of glass recognition techniques is analyzed comprehensively. Next, a glass image segmentation network based on boundary data guidance and the optimization of a planning map based on depth repair are proposed. Finally, map optimization and path-planning tests are conducted and compared using different algorithms. The results confirm the favorable adaptability of the proposed method to indoor transparent plates and glass curtain walls. Using the proposed method, the recognition accuracy of a public test set increases to 94.1%. After adding the planning map, incorrect coverage redundancies for two test scenes reduce by 59.84% and 55.7%. Herein, a glass recognition and map optimization method is proposed that offers sufficient capacity in perceiving indoor glass materials and recognizing indoor no-entry regions.

Keywords Autonomous mobile robot, Multi-sensor fusion, Glass recognition, Map optimization

1 Introduction

The extensive application of autonomous mobile-robot products has resulted in the rapid iteration and development of autonomous navigation and related sensing, localization, mapping, and planning technologies. In this regard, the typical applications include commercial cleaning, disinfection, hotel services, and delivery robots [1]. The primary techniques used are computer vision, point cloud processing, simultaneous localization and mapping (SLAM), motion planning, and multi-robot cooperation. Unlike the structured environment of a laboratory, the dynamic, variable illumination, low-texture, and semi-transparent environmental factors in actual scenarios increase the demand for algorithms with higher robustness and reliability. To satisfy commercial and domestic requirements, some intelligent products, such as autonomous mobile robots, generally adopt low-cost lasers and cameras because of cost restrictions, in addition to ultrasonic, infrared, and inertial navigation systems.

The aforementioned environmental perception schemes satisfy the application requirements of most mobile robots. However, when indoor or outdoor glass curtain walls, glass doors, or transparent dummy plates are present in the environment, the perception of materials may be unsatisfactory [2, 3]. The established planning map may have large gaps and exhibit many potential safety hazards, including risky crossings, collisions, and

*Correspondence:
Yong Tao
Taoy@buaa.edu.cn
[1] School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China
[2] Research Institute of Aero-Engine, Beihang University, Beijing 102206, China
[3] Large Aircraft Advanced Training Center, Beihang University, Beijing 100191, China

falls. For mobile robots, the failure to perceive indoor glass materials and the optimization of planning maps have garnered extensive attention from researchers of robotics and computer vision.

Engineers generally change the operating environment or artificially fix the maps in actual applications to address the problems of glass perception failure and incomplete planning maps. Several approaches have been proposed, such as setting a physical fence before glass screen walls as well as implementing active virtual walls, software visual walls, and magnetic boundaries. However, the failure of glass perception cannot be fundamentally resolved. Owing to the development of intelligent sensors and platform calculation capacities in recent years, multisensor fusion, novel sensors, and new perception algorithms have been proposed to address the abovementioned issue.

Existing SLAM schemes with glass or highly reflective material detection functions primarily use ultrasonic sensors and the distribution of the reflection laser strength as the input. Glass recognition and possibility estimation can be achieved using threshold-limiting and clustering analyses. Despite their low computational requirements and favorable real-time performance, ultrasonic fusion schemes present several problems in practical applications. Strength information is easily affected by multiple factors, including the incident angle [4], detection range, material properties, and platform stability. In addition, a detection window phase exists that depends significantly on the scanning range of LiDAR and a detailed scanning path. Strength reliability is easily affected by glass materials or surrounding diffuse objects of low precision. All these restrictions impose restrictions on the application performance.

By considering low-cost laser and camera perception schemes typically used in mobile robots, a distributed network is presented to further discover visual information without requiring additional sensors. This enables robust glass perception when combined with a laser sensor. Finally, it achieves the perception of glass materials by approximating concrete walls.

This study focuses on an actual mobile robot and proposes an indoor glass recognition and map optimization method based on boundary guidance. First, the interaction mechanisms among different sensors, glass materials, and existing schemes are analyzed. Subsequently, based on typical visual features, a glass-image segmentation network based on boundary guidance is proposed. Next, a planning map optimization method is proposed based on the glass recognition results. Finally, image segmentation, planning map optimization, and planning tests are performed to validate the effectiveness of the proposed method.

The main contributions of this study are as follows: First, the glass image segmentation network is improved based on boundary guidance. Different levels of network structures comprising different modules are investigated to achieve accurate glass image segmentation. Second, a map optimization method based on depth repair is proposed. A glass image segmentation and map optimization scheme is presented based on the integration of LiDAR and a camera. Finally, map optimization and regional full-coverage planning tests are performed on an actual mobile robot to validate the improvement in the performance of the method in enhancing map integrity and planning safety.

2 Related Studies
Existing glass recognition solutions. To address the failure of glass recognition, researchers have considered no-entry or private labeling [5, 6] and changing the operating environment or map modification scheme. This includes the installation of a physical fence, active virtual walls, magnetic boundary lines, adding a software virtual wall, implementing additions, and considering cost map repair in actual engineering applications [7]. Accordingly, the deficiencies of the planning map can be addressed to ensure safe planning. These measures are reliable but not flexible. Virtual software walls have gradually become the standard for service-type mobile robots. Meanwhile, intuitive interactive systems with strict restrictions on vehicle-mounted interactive capabilities are few. Currently, interactive virtual wall operations can be enhanced using mobile devices, laser pens [8, 9], and augmented reality [10]. Although the abovementioned glass recognition schemes based on man–machine interactions have been applied in some cases, autonomous recognition methods for dangerous scenarios, such as those involving glass and stairs, remain challenging.

Glass recognition methods based on LiDAR. LiDAR sensors are typically used in mobile robots and have garnered the attention of researchers. Distance, strength, and echo are the basic combinations of information provided by LiDAR. To address the failure of LiDAR in detecting mirrored objects, Lee et al. [11] employed many single-line LiDAR sensors and used distance data after preprocessing and clustering analysis. They proposed feature extraction methods for glass lines based only on statistical characteristics and coordinate relationships. Jiang et al. [12] proposed an improvement to and a calculation of an occupied grid map based on the confidence degree of glass, as well as calculated the recognition possibility using a neural network classifier. Koch et al. [13] proposed a method for distinguishing mirror reflections from transparent objects based on multi-echo difference and experimentally validated the principal

disadvantage of LiDAR in glass detection. Because multi-echo LiDAR was used, the typical low-cost LiDAR could not satisfy the requirements. Considering the limitation of the incident angle in glass detection by LiDAR, Foster et al. [14] modified the occupied grid map used in the standard SLAM algorithm to survey and map objects visible only at certain view angles by monitoring reliable detectable angle subsets of the objects. Wang et al. [15] proposed a convenient scheme for glass detection in an almost normal range with reflective laser strength. They analyzed the distribution of the reflective laser strength of glass based on the incident angle and screened out the characteristic thresholds of the strength and distribution slope. However, to achieve a stable mirror reflection method, the robot used propagated only along a specified path to scan the suspected glass region along the normal direction of the mirror object surface. This indicates significant restrictions in its application.

Glass recognition methods based on multisensor fusion. Detecting glass using LiDAR alone is difficult. Hence, multisensor fusion schemes have garnered significant attention, which primarily include fusion between a computer vision module comprising a monocular camera, a depth camera, a polarization camera, and LiDAR; fusion between a sonar probe and LiDAR; and fusion between a camera and solar probe. Yang et al. [2] introduced a SLAM fusion method based on 16 sonar probes and a laser scanner to achieve glass detection and a supplementary map. However, the proposed method requires the extraction and reconstruction of disjoint line segments to generate mirror prediction. Additionally, the interference among multiple groups of ultrasonic data reduces the detection precision [16]. Yamaguchi et al. [17] fused a polarization camera and LiDAR to overcome the limitation wherein LiDAR can only detect glass in the normal direction. The initial localization was obtained using LiDAR, and the suspected region was confirmed based on the polarization degree measured by the polarization camera. Using this method, additional polarization cameras are introduced at a high cost for actual applications. Huang and Wei [18, 19] adopted a fusion scheme comprising a camera and two ultrasonic sensors to determine whether glass barriers exist in each frame of an image. Next, the position of the glass was determined using a region-growing algorithm to restore the sparse map. The proposed method accurately detected the position and distance of a glass object in front of a robot with a certain reference significance.

Computer vision-based glass recognition methods Owing to the rapid development of machine learning and the significant enhancement in central processing unit (CPU) and graphics processing unit (GPU) calculation capacity, computer vision technology has been extensively applied in object recognition. The combination of visual cameras and LiDAR is a typical sensor configuration in existing mobile robot platforms that has inspired researchers to combine computer vision with mobile robot sensing schemes. Xie et al. [20] created the Tran10k dataset and proposed the Translab image segmentation method, which has been further developed subsequently. He et al. [21] proposed a glass image segmentation network, i.e., EBLNet, based on edge-learning enhancement. The network first extracts marginal information, followed by scene information within the boundary using a neural network, which significantly enhances the glass image segmentation performance of RGB images. Based on the perspective of computer vision, researchers comprehensively investigated easily confused targets such as glass [22, 23], mirrors [24, 25], and camouflage [26], and rendered the glass detection dataset (GDD) public. The proposed GDnet glass detection network can achieve accurate offline recognition and segmentation of glass regions in images depicting actual scenarios.

For a mobile robot, its first step is to assess whether the object to be detected is glass. Next, the specific coordinates of the glass should be provided accurately and efficiently such that the robot can smoothly complete the functional tasks. These tasks include environmental mapping, self-localization, motion planning, and autonomous navigation. Herein, an indoor glass recognition and map optimization method using platform sensors and background computers is proposed for mobile robots operating based on boundary guidance.

3 Glass Recognition and Map Optimization Method

3.1 Overview of Proposed Method

To address the failure of mobile robots in detecting glass materials, a glass recognition and map optimization method based on boundary guidance is proposed. First, a glass-image segmentation network is proposed based on an existing laser-based SLAM scheme. Environmental RGB images, depth images, and inertial measurement unit location data are obtained and stored, whereas the original grid map is established. The EBLCNet can recognize and manage RGB images, identify environmental glass regions, and segment the boundary ranges of glass materials. The depth images are repaired based on the glass-boundary recognition results. Accordingly, the new grid map for planning is updated offline after the supplementation and excludes dangerous objects, such as glass walls, to ensure that the robot operates in safe cleaning areas, as shown in Figure 1.
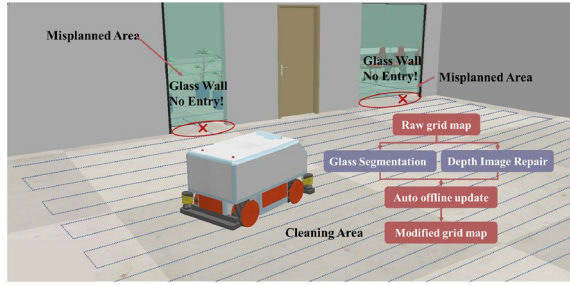
Tao *et al. Chinese Journal of Mechanical Engineering* (2023) 36:74

Page 4 of 12

Figure 1 Diagram showing operation of mobile robot

Figure 2 EBLCNet structure

3.2 Glass Image Segmentation Network Based on Boundary Guidance

3.2.1 Network Structure

Based on the existing object recognition model, the proposed EBLCNet considers the fact that people first focus on the window frame and then the details in the scene. First, the boundary characteristics are extracted using a refined differential module (RDM). Next, a large-field contextual feature integration module is integrated. The scene is recognized and segmented using a boundary feature guidance system. The implementation procedure is described in detail as follows:

The network adopts single RGB images as the input. First, single RGB images are input into a multilayer feature extractor, where a typical pretraining model is used as the backbone. Multiple groups of low- and high-level characteristics are output successively. The RDM module simultaneously receives low- and high-level initial characteristics and calculates the boundary and nonboundary features for the initial learning and iterations. Subsequently, it outputs the combined features and initial boundary characteristics. The LCFI module receives the features and characteristics and uses the initial boundary characteristics as the activation function. It outputs accurate glass-boundary characteristics via guidance training. Subsequently, the loss value is calculated after predictions are obtained using the Pred module and comparing them with the actual values of the dataset. After the iterative loop, binary mask images of the regions labeled as glass are output. Based on a convolutional neural network (CNN), the initial multilevel features are extracted and input to a serial structure comprising an RDM, an LCFI, and Pred for the extraction of continuous characteristics. The serial modules are flexibly arranged to form a multilayer parallel structure. The parallel structures are connected by processed and combined characteristics. The network structure is illustrated in Figure 2.

3.2.2 Fusion of Serial Structure

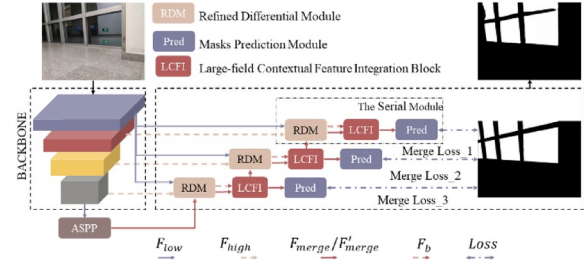After obtaining the initial features and referring to the DeepLabv3+ network structure [27], a serial structure

is set as the main body for feature extraction and processing in the network, as shown in Figure 3. The LCFI module, which includes several LCFI blocks, is adopted after the RDM module, mainly to extract glass features of different sizes. The extracted features improve the adaptability of the segmentation network to glass images at different angles and distances.

Compared with convolution operations with large convolution kernels, spatially isolated convolutions provide more effective, reliable, and feasible schemes of low computational loads for large-range texture extraction. The spatially isolated convolution network is expressed as:

$$F_{out} = \aleph(conv_h(\aleph(conv_v(F_{in})))), \tag{1}$$

where F_{in} and F_{out} are the input and output features of the spatially isolated convolution network, respectively; $conv_h$ and $conv_v$ are convolution operations with $1 \times k$ and $k \times 1$ convolution kernels, respectively; $\aleph$ represents the batch standardization and ReLU calculations.

To eliminate ambiguity and comprehensively extract the features of scene textures, two spatially separated convolutions in two opposite directions are adopted in the LCFI, i.e.,

$$
\begin{aligned}
F_l &= \aleph(conv_1(F_{left\_in})), \\
F_{lcfi} &= \aleph(conv_2(concat(\aleph(conv_v(conv_h(F_l))), \\
&\quad \aleph(conv_h(conv_v(F_l))))))),
\end{aligned} \tag{2}
$$
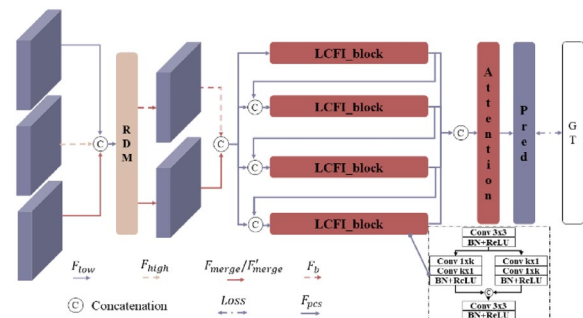
Figure 3 Illustration of EBLCNet serial structure

Tao *et al. Chinese Journal of Mechanical Engineering* (2023) 36:74

Page 5 of 12

where F_{left_in} and F_{lcfi} denote the input and output features of the LCFI module, respectively; $conv_1$ and $conv_2$ represent ordinary convolution; $conv_h$ and $conv_v$ represent convolution operations with $1 \times k$ and $k \times 1$ convolution kernels, respectively; and $\aleph$ represents the batch standardization and ReLU calculations.

In contrast to the single use of the RDM or LCFI module, the fused serial network structure can fully utilize the boundary and hybrid features extracted from the initial features. It recognizes glass doors and windows on a single scale using a single LCFI module. By connecting it with the LCFI module in parallel, a large-scale detection input is adopted as the output in the next module to achieve multiscale recognition of glass doors and windows. After obtaining accurate boundary recognition results, a new round of prediction and iteration is performed for a layer in the serial structure. The output from the current layer of the serial structure is adopted as the input for the next layer of the serial structure.

3.2.3 Loss Function

The loss function is categorized into a single-layer loss function and network-structure loss function. By referring to the EBLNet, the one-way loss function comprising boundary loss, nonboundary loss in the RDM module, and serial hybrid loss is written as:

$$L_{joint} = \lambda_1 L_{residual}(F_r, G_r) + \lambda_2 L_{edge}(F_b, G_e) + \lambda_3 L_{merge}(F_m, G_m), \quad (3)$$

where F_m, F_b, and F_r are the single-layer hybrid, boundary, and nonboundary features in the RDM module, respectively; G_m, G_e, and G_r are the actual values of the

$$L = \sum_{n=1}^{N} L_{joint}^n. \quad (4)$$

The network loss function is calculated to train and optimize the entire network to enhance the accuracy.

3.3 Optimization of Planning Map Based on Depth Repair

Depth image repair. A depth camera provides a robot with RGB and depth images, which can serve as an effective supplement to the distance data provided by LiDAR. However, owing to the limitations of ranging, a significant amount of noise and void defects exist in the glass region of the original depth images. These images are not conducive to the direct sampling and supplementation of the environmental map. Therefore, an optimization method for the planning map is proposed based on depth repair. It repairs voids in depth images using glass image recognition and segmentation. Subsequently, the grid map is automatically supplemented via depth interpolation and sampling to achieve path planning for the mobile robot.

The following two methods are used to repair noisy points and voids in the image. First, for defects of few pixels, the distance can be supplemented using median filtering. Subsequently, for large-area voids, the distance values around the pixels are calculated and supplemented via linear filtering based on the glass image segmentation boundary results. For void defects, assuming that E denotes the set of segmentation boundary points and B denotes the point set, the output depth values are generally zero, and the following relationship applies for any point:

$$P_1^B = \frac{(d_A + d_D)\left(d_W P_1^{E_S} + d_S P_1^{E_W}\right)}{(d_W + d_S + d_A + d_D)(d_W + d_S)} + \frac{(d_W + d_S)\left(d_D P_1^{E_A} + d_A P_1^{E_D}\right)}{(d_W + d_S + d_A + d_D)(d_A + d_D)}, \quad (5)$$
$$P_r = P_1 + P_1^B,$$

image, the boundary value, and the nonboundary value, respectively; L_{merge} and $L_{residual}$ are the layer mixing loss and nonboundary loss in the RDM module, respectively, which adopt a mixed cross-entropy [28]; L_{edge} denotes the boundary loss in the RDM module, which adopts the Dice loss [29]; λ denotes the weight coefficient; and G_e and G_r are screened and extracted from G_m via complementation and residue calculations, respectively.

Considering the connection of serial structures in parallel, the loss function of the network is calculated by adding the losses of N serial layers as follows:

where P_1, P_r, and P_1^B are the initial depth image, depth image after repair, and repair-process values, respectively; d_W, d_S, d_A, and d_D are the shortest Euclidean distances of the defect points from the boundary points in the four directions; and $P_1^{E_W}$, $P_1^{E_S}$, $P_1^{E_A}$, and $P_1^{E_D}$ are the depths of the corresponding boundary points in the four directions.

Sampling supplementation. To supplement the depth data, the distance information should be subjected to dimension reduction. Data supplementation is performed on the original grid map within
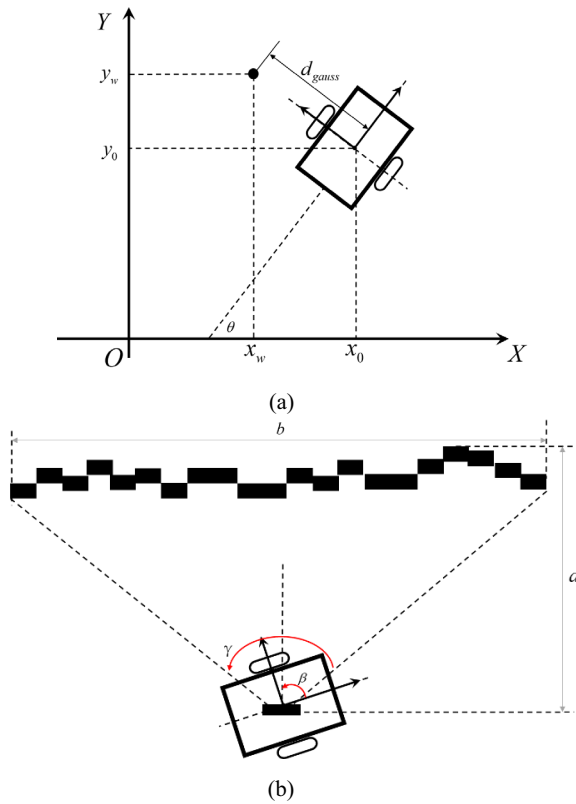
Tao *et al. Chinese Journal of Mechanical Engineering*   (2023) 36:74

Page 6 of 12

Figure 4 Illustration of depth image sampling supplementation: **a** coordinate system establishment, **b** sampling supplementation schematic

Figure 5 Experimental validation and decomposition

measured by LiDAR at the grid point in the view field and camera, respectively; δ is the degree of confidence between two measured values; And θ, β, and γ denote the azimuth angle of the robot, the included angle between the pointing direction and advancing direction of the camera, and the field angle of the camera, respectively.

The archived data are traversed at all grid points on the map boundary to achieve sampling supplementation on the regional map.

Offline update. The updated two-dimensional grid map adequately restores the actual environmental boundaries to update the static map layer with via the cost map mechanism.

4 Experiments

This section describes the experimental scheme, platform, and datasets used in this study. For actual indoor glass scenes, glass image segmentation, planning map optimization, and covering path planning tests were conducted to validate the effectiveness and feasibility of the proposed method.

4.1 Experimental Settings

Figure 5 shows the experimental setup. The present experiment included three main tests: the glass image segmentation test, planning map optimization test, and covering path planning contrast test.

A glass-image segmentation network was established based on the PyTorch framework. Typical ResNet and ResNeXt network modules were used in the backbone multilayer feature extractor network. During the training process, the GDD and mirror detection dataset (MSD) were used. The GDD included 2980 training images and related masks under multiple scenes, such as shopping malls, stations, houses, and offices. The MSD included 3063 training images and related masks for multiple types of mirrors and mirrors obstructed by objects.

The learning rate was set to 0.001–0.003. Consecutive training was conducted under 100 iterations. The computer used for network training featured a Windows 10 operating system, an Intel CPU i5-12400F, and 16G of random access memory. A Tesla T4 and a P100 GPU were used to accelerate the calculations. The calculations required approximately 20 h. A total of 936 GDD and 955 MSD images were randomly selected for validation.

the depth-camera measurement range. The data are directly supplemented at the barrier position in the original direction. When the barrier is along the original direction, the distance difference threshold ε is set. If the difference between the barrier distance in the original map and the data exceeds the threshold, then an object with a small value is regarded as a barrier to ensure the safety of the planned path. Otherwise, the distance data are combined and the new distance d_{gauss} is calculated via Gaussian filtering and then added to the map, as shown in Figure 4.

$$\begin{cases} d_{gauss} = \rho d_{obstacle}^m + \delta d_{camera}^m, \\ x_w = x_0 + d_{gauss} \cos\left(\theta + \beta + \dfrac{\gamma}{2} - \dfrac{\gamma m}{M}\right), \\ y_w = y_0 + d_{gauss} \sin\left(\theta + \beta + \dfrac{\gamma}{2} - \dfrac{\gamma m}{M}\right), \end{cases} \quad (6)$$

where M and m denote the number of grids within the visual field and the related serial number, respectively; x_w, y_w, x_0, and y_0 are the coordinates of the occupied grids and robots in the world coordinate system; d_{gauss} denotes the Euclidean distance between the occupied grid and camera plane; $d_{obstacle}^m$ and d_{camera}^m are the distance values
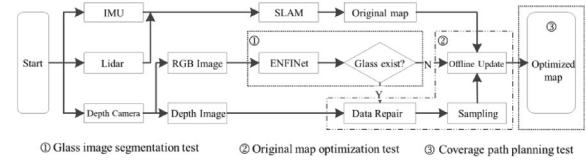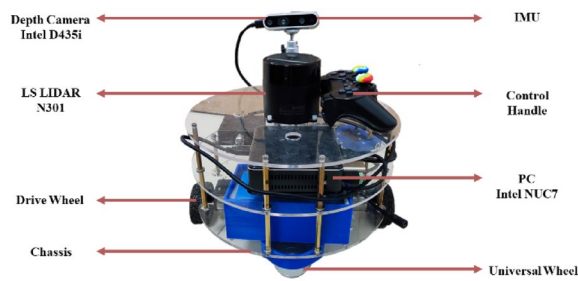
Tao *et al. Chinese Journal of Mechanical Engineering*   (2023) 36:74

Page 7 of 12

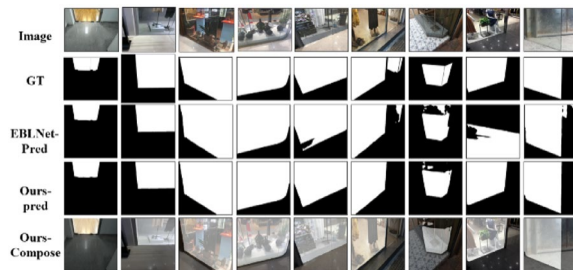**Figure 6** Mobile robot platform

Figure 7 Comparison of image segmentation results

As shown in Figure 6, the self-developed mobile robot implemented environmental mapping and planning after map supplementation. The main components of the robot were a frame, a wheel system, an STM32 control panel, and a power system. The LSLIDAR-N301 LiDAR

system was obtained from Leishen Intelligence. The NUC7i5BNH host machine was obtained from Intel. A D435i depth camera and an IMU were mounted. The robot chassis performed differential driving for advancement and used ROS-Kinetic for communication with the Ubuntu 16.04 operating system.

4.2 Glass Image Segmentation Test

For an unbiased comparison, the parameter settings were almost identical to those of the GDNet and EBLNet. Images with a resolution of 416×416 were input, and the learning rate was set to 0.002 after dynamic regulation using an attenuation parameter of 0.9, based on the Poly strategy. The number of network layers determines the number of parameters affecting the learning speed. The training results of the recognition network with different numbers of serial structures were tested and analyzed. The image segmentation performance improved compared with the results obtained using previous algorithms. In a scenario with abundant window boundary textures, precise segmentation performance is achieved to supplement the planning map. Figure 7 shows a visual comparison of the image segmentation results yielded by the different algorithms.

To evaluate the efficiency and accuracy, the results obtained using similar algorithms in the domain were added for comparison. The advantages and enhancements of the EBLCNet in terms of both accuracy and flexibility were observed. In this study, several indices, including the intersection of union (IoU), accuracy (Acc),

Table 1 Comparison of test results for GDD

Method	CRF	IoU↑	Acc↑	F-beta↑	mAE↓	BER↓
PSPNet	−	84.06	0.906	0.906	0.084	8.79
DenseASPP	−	83.68	0.919	0.911	0.081	8.66
DANet	−	84.15	0.911	0.901	0.089	8.96
CCNet	−	84.29	0.915	0.904	0.085	8.63
PointRend	−	86.51	0.933	0.928	0.067	6.50
DSS	−	80.24	0.898	0.890	0.123	9.73
PiCANet	−	83.73	0.916	0.909	0.093	8.26
BASNet	−	82.88	0.907	0.896	0.094	8.70
EGNet	−	85.04	0.920	0.916	0.083	7.43
DSC	−	83.56	0.914	0.911	0.090	7.97
BDRAR	+	80.01	0.902	0.908	0.098	9.87
GDNet (RNeXt101)	+	87.63	0.939	0.937	0.063	5.62
EBLNet (RNet101)	−	88.16	0.941	0.939	0.059	5.58
EBLNet (RNeXt101)	−	88.72	0.944	0.940	0.055	5.36
EBLCNet (ResNet101) 1C	−	88.38	0.939	0.945	0.061	6.01
EBLCNet (ResNet101) 2C	−	88.71	0.941	0.941	0.059	5.76
EBLCNet (ResNet101) 3C	−	88.68	0.941	0.941	0.059	5.86

Conditional random field (CRF) is used for post-processing

Table 2 Comparison of test results based on MSD

Method	CRF	IoU↑	Acc↑	F-beta↑	mAE↓	BER↓
MirrorNet	+	78.95	0.935	0.857	0.065	6.39
EBLNet (RNet101)	−	78.84	0.946	0.876	0.054	8.84
EBLNet (RNeXt101)	−	80.33	0.951	0.883	0.049	8.63
EBLCNet (RNet101) 2C	−	83.26	0.934	0.929	0.066	10.11
EBLCNet (RNeXt101) 2C	−	84.69	0.940	0.938	0.059	9.44

CRF is used for post-processing

mean absolute error (mAE), and balanced error rate (BER), were calculated for comparison. The results are summarized in Table 1.

Tables 1 and 2 present a comparison of the test results, where C denotes the number of serial structure layers. The proposed two-layer serial structures achieve the training performance of three layers using a more convenient network and less data. EBLCNet fuses boundary perception and a large reception field convolution module. By performing training and validation based on the GDD and MSD, the accuracy of recognizing glass doors and windows and image segmentation further improved. Compared with EBLNet, EBLCNet can manage more data. However, this difference becomes less prominent after the layers of the serial structures are adjusted. Considering the environmental map and images obtained by the mobile robot, the recognition task was deployed in the background for offline operation, which did not impose a high demand on real-time capability. In tests using the hardware above based on public datasets, the inference time of the proposed method was approximately 65 ms, which was slightly longer than those of EBLNet (24 ms) and GDNet (41 ms) using ResNet50 as the backbone.

4.3 Original Map Optimization Test

To deploy the recognition algorithm and validate the performance of the proposed optimization method for automatic mapping, an actual office environment obstructed by a glass wall was selected as the operating environment for the mobile robot. A two-dimensional navigation map was established using the Gmapping algorithm [30]. During the mapping process, RGB images, depth images, and the corresponding IMU location data were recorded. Subsequently, they were unloaded to the server for an offline recognition of unknown glass regions. Next, depth images were sampled to complete the automatic supplementation of the planning map. Figure 8 shows the mapping process during the movement of the robot.


Figure 8 Mapping process of robot

Figure 9 shows the original mapping results and environmental characteristics. (A) and (C) show the enlarged details of the glass doors and windows in the test environment, respectively.

Establishing an environmental map using LiDAR is applicable to most scenes, such as those involving walls; however, some glass doors, windows, and partition regions cannot be perceived. For glass windows, LiDAR can only detect the existence of a window frame, as shown in the enlarged images (B) and (D) in Figure 9. The established map incorrectly includes most of the

Tao *et al. Chinese Journal of Mechanical Engineering*   (2023) 36:74

Page 9 of 12
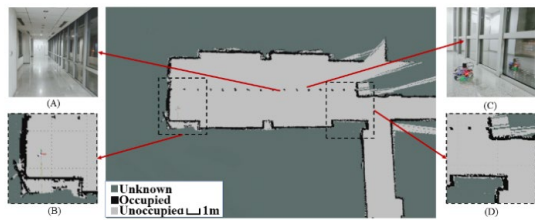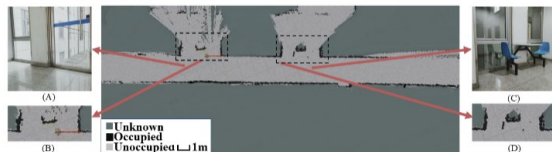
(a) Scene 1

(b) Scene 2

Figure 9 Mapping test scene in office area
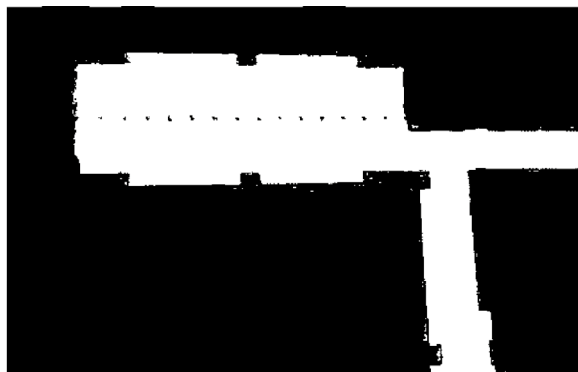
(a) Original mapping result of scene 1

(b) Optimized mapping result of scene 1

Figure 10 Comparison of map **a** before and **b** after supplementation for Scene 1

(a) Original mapping result of scene 2

(b) Optimized mapping result of scene 2

Figure 11 Comparison of map **a** before and **b** after supplementation for Scene 2

This results in a high risk of collision and falling during the implementation of typical tasks such as fixed-point distribution and regional cleaning.

In Figures 10 and 11, the occupied grids are marked in black. The passable grids are marked in white. Figures 10(a) and 11(a) show the results of the original two-dimensional laser mapping. In the original map, the metal frames of the doors and windows were detected and marked in black. However, glass was not detected and was incorrectly marked as white. Figure 10(b) and 11(b) show the results of the supplementary and updated maps. After supplementing and updating the planning map, the glass region was relabeled as a no-entry state.

4.4 Coverage Path Planning Test

To validate the contribution of the map supplementation to the safety of autonomous navigation by the robot, a full-coverage planning test was conducted to simulate the cleaning task. The aim of this test is to achieve regional coverage. In Figure 12, the green pentagram and red triangle represent the starting point of planning and the end of the operation, respectively. The bow-type yellow line represents the path coverage.

For the full-coverage planning for Scenes 1 and 2, the original planning map for the mobile robot misidentified the unknown region from the glass door and rear region. The end point was located outside the glass door, and the route obtained by the planning algorithm passed through the glassy region. During actual operation, the
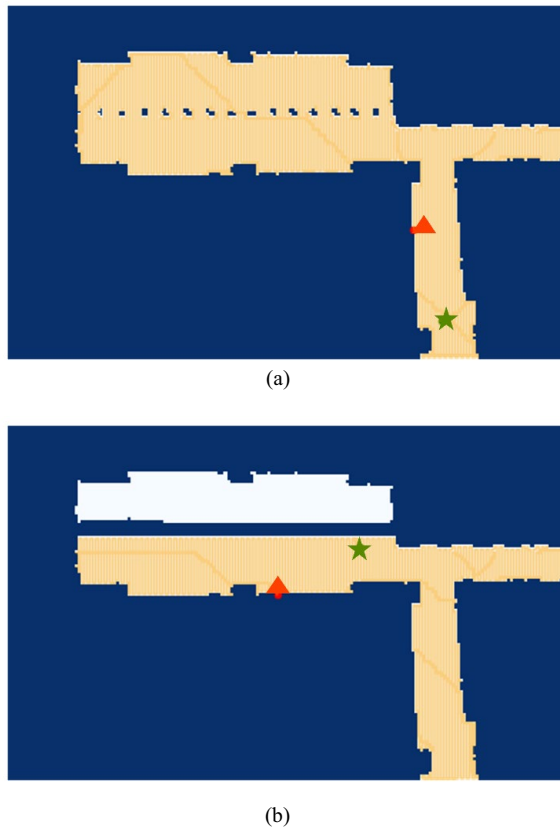
unknown regions. Without map repair and adjustment, the robot considers the entire domain in the map as safe and passable, based on the planning algorithm.

Figure 12 Comparison of coverage planning results for using **a** original and **b** optimized map of Scene 1

Figure 13 Comparison of coverage planning results using **a** original and **b** optimized map of Scene 2

robot system may collide or fall. After supplementing the planning map for the full-coverage simulation, the entire path was within the corridor without risk objects, such as glass and stairs. Hence, the operational safety of the robot was ensured, as shown in Figures 12 and 13.

To evaluate the improvement in the map supplementation quantitatively, five indices were selected to measure the difference between the planning paths in different phases and in the building map. The indices were the grid occupation rate (GOR), obstacle point number (OPN), coverage path length (CPL), incorrect coverage redundancy (WCR), and collision possibility (CP). Table 3 lists the detailed data before and after map optimization.

As shown in Table 3, owing to the supplementation of the boundary information, the GORs in the Scenes 1 and 2 enhanced by 1.13% and 2.03%, respectively. After supplementation, the planning map segmented and avoided the unreliable glass boundary and rear regions in the original map, thereby reducing the WCRs by 59.84% and 55.7% in Scenes 1 and 2, respectively. In general, both the accuracy and efficiency of the path planning were effectively enhanced in the cleaning task simulation.
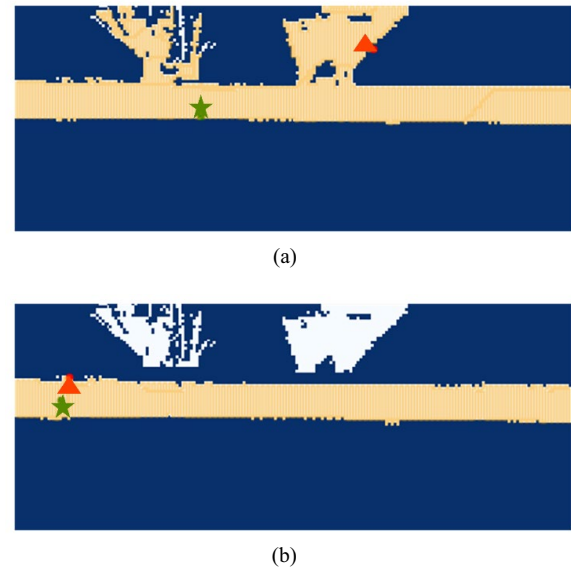
4.5 Discussions
Segmentation network. Without the addition of vehicle sensors, the proposed method relies on low-cost LiDAR and depth cameras and achieves glass-region segmentation based on RGB images via a glass-image segmentation network. In particular, the proposed method combines glass recognition with the optimization of the robot's planning map. Based on the GDD, MSD, and previous studies, different backbone networks could be constructed and tested. Finally, the proposed method achieves an accuracy rate of 94.1% for glass recognition and segmentation.

By considering the boundary data and large-reception-field convolution module, the established network can be applied to many target scenes, including glass doors and windows. However, during the actual training process, the established network could not readily recognize the intersecting regions among the floor bricks, skirting lines, and glass. The indoor arch door could not be distinguished easily from the boundless door by relying only on the boundary information for guidance; thus, the further enhancement of the recognition accuracy was restricted.

Autonomous navigation. Based on the principles of online acquisition, offline recognition, and automatic supplementation, the robot performed data acquisition and storage during the first operation. First, the image data required for recognition and the IMU data were stored. Next, offline image recognition was conducted based on the pretrained model in free time. Finally, the results were automatically updated based on the recognition results and update rules. In essence, the proposed

Tao *et al. Chinese Journal of Mechanical Engineering*   (2023) 36:74

Page 11 of 12

Table 3 Comparison of evaluation indexes before and after supplementation

		GOR (%)	OPN	CPL	WCR (%)	CP
Scene 1	Original map	75.72	12266	4740	61.11	+
	Optimized map	76.85	12450	2942	1.27	−
Scene 2	Original map	74.85	19012	7152	56.84	+
	Optimized map	76.88	19527	4560	1.14	−

method updates only the static map layer and restricts the planning of safe regions. Thus, the WCR is reduced and the operational safety of the robot is ensured. However, the location accuracy of the mobile robot did not indicate much improvement. The constant improvement in hardware calculation capacity necessitates further improvement in low computational cost networks, online recognition, multisensor fusion localization, and multimap management schemes.

Influence of lights. In this study, the public dataset and our self-established image test set accounted for various scenarios such as homes, shopping malls, and offices. Currently, they are operated under good lighting conditions. The proposed method relies on the features of the image, and the improvement in recognition accuracy is mainly due to the deep mining of glass boundary features. Therefore, in principle, this method is only applicable under well-lit conditions. Under conditions without any lights and with direct sunlight, intense reflections or light spots are presented on the glass doors and windows. However, this scenario is challenging. Fusion schemes that include polarizing cameras or other heterogeneous sensors are possible solutions to the abovementioned issue.

5 Conclusions

(1) To address the failure of mobile service robots in perceiving indoor glass materials, the status of studies pertaining to glass recognition techniques was analyzed in terms of algorithms and products. Difficulty in detecting special objects, such as glass, hinders the application of autonomous mobile robots.

(2) A glass recognition and map optimization method based on boundary guidance was proposed. The proposed method involved a glass image segmentation network and a map optimization algorithm.

(3) Map optimization and planning tests were performed using different algorithms. The proposed method exhibited favorable adaptability to indoor transparent plates and glass curtain walls. The rec-

ognition accuracy of the public test set increased to 94.1%. After supplementing the planning map, the WCRs of the Scenes 1 and 2 reduced by 59.84% and 55.7%, respectively.

(4) Deploying the proposed method to actual robots can simplify certain operations, such as the artificial labeling of glass regions and the addition of virtual walls during the mapping process. Thus, the integrity of the planning map and the safety of the path-planning process are ensured.

Authors' Contributions
YT was in charge of the entire trial, HG wrote and reviewed the manuscript, and LD checked the paper format and assisted in setting up the hardware platform. YW and JL assisted with the data collection and laboratory analyses. All authors read and approved the final manuscript.

Authors' Information
Yong Tao was born in 1979 and is currently an associate professor at *Beihang University, China*. He received his Ph.D. degree from *Beihang Universtiy, China*, in 2009. His research interests include mechanic engineering and robotics.
He Gao, born in 1997, is currently a Ph.D. candidate at *Research Institute of Aero-Engine, Beihang University, China*. His research interests include mobile manipulations and motion planning.
Yufang Wen was born in 1997. He is currently pursuing a master degree at *School of Mechanical Engineering and Automation, Beihang University, China*. His research interests include feature detection, motion planning, and tracking control for autonomous mobile robots.
Lian Duan, born in 1996, is currently pursuing a master degree at *Large Aircraft Advanced Training Center, Beihang University, China*. His research interests include intelligent aerospace manufacturing and motion planning.
Jiangbo Lan was born in 1996 and is currently pursuing a master degree at *Large Aircraft Advanced Training Center, Beihang University, China*. His research interests include intelligent control, manufacturing and robotic manipulation.

Declarations

Competing Interests
The authors declare no competing financial interests.

Tao *et al. Chinese Journal of Mechanical Engineering*   (2023) 36:74

Page 12 of 12

References

[1] M Cardona, F Cortez, A Palacios, et al. Mobile robots application against covid-19 pandemic. *IEEE ANDESCON,* Quito, Ecuador, Oct 13-16 2020: 1-5.

[2] S W Yang, C C Wang. Dealing with laser scanner failure: mirrors and windows. *IEEE International Conference on Robotics and Automation (ICRA),* California, USA, May 19-23, 2008: 3009-3015.

[3] J Lu. *Research on safety coping strategies in application of building glass curtain wall.* Harbin: Harbin Institute of Technology, 2018.

[4] G Chartier. *Introduction to optics.* New York: Springer, 2005.

[5] B Gromov, G Abbate, L M Gambardella, et al. Proximity human-robot interaction using pointing gestures and a wrist-mounted IMU. *IEEE International Conference on Robotics and Automation (ICRA),* Montreal, Canada, May 20-24, 2019: 8084-8091.

[6] N Hawes, C Burbridge, F Jovan, et al. The STRANDS project: long-term autonomy in everyday environments. *IEEE Robotics Automation Magazine,* 2017, 24(3): 146-156.

[7] W X Liu. *Design and implementation of control and management system of indoor sweeping robot.* Beijing: Beijing University of Posts and Telecommunications, 2021.

[8] D Sprute, K Tönnies, M König. This far, no further: Introducing virtual borders to mobile robots using a laser pointer. *IEEE International Conference on Robotic Computing (IRC),* Naples, Italy, June 16-19, 2019: 403-408.

[9] D Sprute, K D Tnnies, M Koenig. Interactive restriction of a mobile robot's workspace in a smart home environment. *Journal of Ambient Intelligence and Smart Environments,* 2019, 11(6): 475-494.

[10] D Sprute, K D Tnnies, M Koenig, et al. Virtual borders: Accurate definition of a mobile robot's workspace using augmented reality. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Madrid, Spain, October 1-5, 2018: 8574-8581.

[11] S H Lee, J H Oh, Y C An. A new range-only measurement-based glass line feature extraction method. *Electronics Letters,* 2021, 57(21): 804-806.

[12] J Jiang, R Miyagusuku, A Yamashita, et al. Online glass confidence map building using laser rangefinder for mobile robots. *Advanced Robotics,* 2020, 34(23): 1506-1521.

[13] R Koch, S May, A Nuchter. Effective distinction of transparent and specular reflective objects in point clouds of a multi-echo laser scanner. *IEEE International Conference on Advanced Robotics (ICAR),* Hong Kong, China, May 29 to June 3, 2017: 566-571.

[14] P Foster, Z Sun, J J Park, et al. VisAGGE: visible angle grid for glass environments. *IEEE International Conference on Robotics and Automation (ICRA),* Karlsruhe, Germany, May 6-10, 2013: 2213-2220.

[15] X Wang, J Wang. Detecting glass in simultaneous localisation and mapping. *Robotics & Autonomous Systems,* 2017, 88: 97-103.

[16] J Su, J Li, Y Zhang, et al. Selectivity or invariance: Boundary-aware salient object detection. *IEEE/CVF International Conference on Computer Vision (ICCV),* Seoul, South Korea, October 27 to November 2, 2019: 3799-3808.

[17] E Yamaguchi, H Higuchi, A Yamashita, et al. Glass detection using polarization camera and LRF for SLAM in environment with glass. *International Conference on Research and Education in Mechatronics (REM),* Cracow, Poland, December 9-11, 2020: 1-6.

[18] Z Huang, K Wang, K Yang, et al. Glass detection and recognition based on the fusion of ultrasonic sensor and RGB-D sensor for the visually impaired. *SPIE Target and Background Signatures,* Berlin, Germany, September 10-13, 2018: 118-125.

[19] H Wei, X E Li, Y Shi, et al. Multi-sensor fusion glass detection for robot navigation and mapping. *WRC Symposium on Advanced Robotics and Automation (WRC SARA),* Beijing, China, August 18-21, 2018: 184-188.

[20] E Z Xie, W J Wang, W H Wang, et al. Segmenting transparent objects in the wild. *European Conference on Computer Vision (ECCV),* Glasgow, UK, August 23–28, 2020: 696-711.

[21] H He, X Li, G Cheng, et al. Enhanced boundary learning for glass-like object segmentation. *IEEE/CVF International Conference on Computer Vision (ICCV),* Montreal, Canada, March 10, 2021: 15859-15868.

[22] H Mei, X Yang, Y Wang, et al. Don't hit me! Glass detection in real-world scenes. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* WA, USA, June 16-18, 2020: 3687-3696.

[23] Y Li and M S Brown. Single imagelayer separation using relative smoothness. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Ohio, USA, June 23-28, 2014: 2752-2759.

[24] X Yang, H Mei, K Xu, et al. Where is my mirror? *IEEE/CVF International Conference on Computer Vision (ICCV),* Seoul, Korea (South), June 24-27, 2019: 8809-8818.

[25] H Mei, B Dong, W Dong, et al. Depth-aware mirror segmentation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* June 19-25, 2021: 3044-3053.

[26] H Mei, Y Liu, Z Wei, et al. Exploring dense context for salient object detection. *IEEE Transactions on Circuits and Systems for Video Technology,* 2021, 32(3): 1378-1389.

[27] L C Chen, Y Zhu, G Papandreou, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European Conference on Computer Vision (ECCV),* Munich, Germany, Sept 8-14, 2018: 801-818.

[28] H Li, W Lu. Mixed cross entropy loss for neural machine translation: Proceedings of machine learning research. *International Conference on Mechine Learning,* 2021: 6425-6436.

[29] F Milletari, N Navab, S Ahmadi. V-Net: fully convolutional neural networks for volumetric medical image segmentation. *International Conference on 3D Vision (3DV),* CA, USA, October 25-28, 2016: 565-571.

[30] G Grisetti, C Stachniss, W Burgard. Improved techniques for grid mapping with Rao-Blackwellized particle Filters. *IEEE Transactions on Robotics,* 2007, 23(1): 34-46.