3-D-Laser-Based Scene Measurement and Place Recognition for Mobile Robots in Dynamic Indoor Environments

Yan Zhuang, Nan Jiang, Huosheng Hu, *Senior Member, IEEE*, and Fei Yan

***Abstract*—Active environment perception and autonomous place recognition play a key role for mobile robots to operate within a cluttered indoor environment with dynamic changes. This paper presents a 3-D-laser-based scene measurement technique and a novel place recognition method to deal with the random disturbances caused by unexpected movements of people and other objects. The proposed approach can extract and match the Speeded-Up Robust Features (SURFs) from bearing-angle images generated by a self-built rotating 3-D laser scanner. It can cope with the irregular disturbance of moving objects and the problem of observing-location changes of the laser scanner. Both global metric information and local SURF features are extracted from 3-D laser point clouds and 2-D bearing-angle images, respectively. A large-scale indoor environment with over 1600 m^2 and 30 offices is selected as a testing site, and a mobile robot, i.e., SmartROB2, is deployed for conducting experiments. Experimental results show that the proposed 3-D-laser-based scene measurement technique and place recognition approach are effective and provide robust performance of place recognition in a dynamic indoor environment.**

***Index Terms*—Autonomous place recognition, bearing-angle image, Speeded-Up Robust Features (SURFs), 3-D laser scanning, 3-D-laser-based scene measurement.**

I. Introduction

AUTONOMOUS scene measurement and place recognition are closely related to navigation, localization, and mapping tasks of mobile robots. Recently, a variety of place recognition algorithms have been developed for mobile robots to implement place recognition in diversified indoor environments, and a rich body of research results has been published in the robotics literature [1]–[5]. Most of these approaches are based on passive vision sensors.

Quattoni and Torralba proposed a monocular-vision-based place recognition model that is specifically tailored to the task of indoor place recognition [1]. It successfully combines local and global discriminative information. Cummins and Newman proposed the feature appearance-based mapping algorithm [2], in which omnidirectional camera images were used. A lightweight descriptor for omnidirectional vision was presented in [3], which is invariant to rotation and slight changes of illumination. In order to handle orientation changes of a mobile robot traversing the same area, a global gist descriptor was utilized in [4] to capture the basic structure of different types of scenes in a very compact way. Cadena *et al.* presented a stereo vision system based on a bag-of-words algorithm and conditional random fields, in order to solve the place recognition problem, and achieved more robust results than using spatial consistency based on epipolar geometry [5]. Although passive vision sensors are very powerful and widely used in mobile-robot navigation, they cannot work in a dark environment and are very sensitive to illumination changes.

On the other hand, laser scanning sensors have a field of view wider than most monocular or stereo cameras, and their data are able to provide detailed range information for accurate place recognition by mobile robots. A great diversity of range sensors has been developed for mobile robots to acquire range date from their environments, including 2-D laser range finders (e.g., SICK or Hokuyo), the SR4000 Swiss Ranger, all-solid-state time-of-flight 3-D camera PMD CamCube 3.0, and the low-cost 2-D range camera systems from Canesta and Primesense [16]. Ibeo LUX laser scanner has a field of view up to 200 m and accurately and reliably works at high speeds, poor weather conditions, and heavy traffic. It has been used in urban traffic and motorway. However, the vertical field of view is 3.2°, which cannot meet the demand of the indoor place recognition task. Moreover, Leica HDS3000 terrestrial laser scanner is able to collect point clouds for registration of large range data sets, which are more precise than those obtained from SICK LMS 291 in our paper. Velodyne HDL-64E LiDAR sensor has 360° field of view and the high data rate that makes it ideal for the most demanding perception applications, as well as 3-D mobile data collection and mapping applications. It was also the primary means of terrain map construction and obstacle detection for many top Defense Advanced Research Projects Agency Urban Challenge teams. However, Leica HDS 3000 and Velodyne HDL-64E are expensive for mobile-robot indoor scene measurement and place recognition.

In recent years, a lot of research has been focused on 3-D laser measurement, 3-D laser mapping, and so on [6]–[10]. Magnusson *et al.* proposed a novel place recognition approach

Manuscript received April 23, 2012; revised July 4, 2012; accepted August 1, 2012. Date of publication September 18, 2012; date of current version December 29, 2012. This work was supported in part by the National Natural Science Foundation of China under Grant 61075094 and Grant 61035005 and in part by the Fundamental Research Funds for the Central Universities of China under Grant DUT11ZD201. The Associate Editor coordinating the review process for this paper was Dr. Shervin Shirmohammadi.

Y. Zhuang and N. Jiang are with the Research Center of Information and Control, Dalian University of Technology, Dalian 116024, China.

H. Hu is with the School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, U.K.

F. Yan is with the Department of Electrical Engineering, The City College, City University of New York, New York, NY 10031 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2012.2216475

in which the normal distributions transform was used as a global description of the appearance of a 3-D scan [11]. Another place recognition approach based on a novel feature descriptor was proposed in [12], in which range images were considered as a representation of laser scans and used for point feature extraction. The method proposed in [11] has substantially lower recall rates, and their feature descriptor is not invariant to changes in the roll and the pitch of an object [13]. Spin images are a typical feature to represent local 3-D point clouds for surface matching and object recognition [14]. However, 3-D point clouds need high measuring accuracy for matching scenes in spin images, resulting in a high computational cost. Therefore, spin images are not a practical choice for real-time place recognition in this paper.

The aim of this paper is to design a 3-D-laser-based place recognition system for a mobile robot to recognize complex indoor scenes autonomously, and to deal with the disturbance from moving objects and people effectively. First, a 3-D-laser-based scene measurement technique and the corresponding scene feature description approach are proposed for indoor place recognition, which consists of both the local Speeded-Up Robust Features (SURFs) extracted from bearing-angle images and the global spatial features extracted from raw 3-D point clouds. Then, a novel place recognition framework is proposed, which has a large and diverse database for handling indoor place recognition. Based on our previous work in [15], this paper deploys SURF features instead of scale-invariant feature transform (SIFT) features, so that the computational cost for local feature extraction and place recognition can be greatly reduced. Moreover, a door detection algorithm and a place recognition guideline extraction algorithm are developed such that a mobile robot can perform indoor place recognition robustly. To show the validity and the efficiency of the proposed approach, an extensive experimental evaluation is conducted.

The rest of this paper is organized as follows: Section II briefly describes the process of 3-D laser data acquisition and their transformation to bearing-angle images. Section III presents novel algorithms used for SURF features extraction and matching based on bearing-angle images, including how to extract the global spatial information. The implementation of indoor place recognition by calculating and scoring the candidate transformation is introduced in Section IV. Experimental results and analysis are presented in Section V to demonstrate the feasibility and the effectiveness of the proposed approach. Finally, a brief concluding remark and future work are given in Section VI.

II. 3-D Laser Data Acquisition and Bearing-Angle Image Generating

A. 3-D Laser Data Acquisition

In this paper, the proposed 3-D-laser-based scene measurement technique is implemented on a self-built data acquisition system, which is realized by rotating a 2-D laser range finder (SICK LMS 291, 180° scan with 0.5° resolution) on a rotating platform, as shown in Fig. 1(a). This built-in-house 3-D laser scanner had been placed on a SmartROB2 mobile robot and tested in a series of indoor scenes. In our paper, the robot uses a 3-D laser range finder as the only sensor to perceive environment information, and it takes about 4 s to capture a scan, which consists of 52 000–56 000 points.

The laser point clouds obtained from the 3-D laser scanner can be represented in polar coordinates (ρ, θ, φ), where ρ is the distance from the optical center of the laser range finder to the detected object, θ is the angle in the scanning plane, and φ is the elevation of the rotator [see Fig. 1(b)]. These polar coordinates are also defined as the robot coordinates. The physical meaning of parameters a, b, and c are the same with those in Fig. 1(c). The 3-D laser scanning data in the local Cartesian coordinate system (x, y, z) can be calculated by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \sin\theta & 0 \\ \sin\varphi\cos\theta & \cos\varphi \\ -\cos\varphi\cos\theta & \sin\varphi \end{bmatrix} \begin{bmatrix} \rho \\ b \end{bmatrix} + \begin{bmatrix} 0 \\ c\cdot\sin\varphi \\ a - c\cdot\cos\varphi \end{bmatrix}. \tag{1}$$

B. Bearing-Angle Image Generating

Range images are a conventional pictorial representation of 3-D laser scanning data. In order to compare with bearing-angle images, the range images are displayed according to the format of data storage, with depth values placed in 2-D arrays. Fig. 2(b) shows the range image with gray shade proportional to the depth of the scene depicted by the raw 3-D laser scanning point cloud in Fig. 2(a), where the 3-D laser scanning covers an elevation of 50°. Although the depth discontinuities of the scene are shown in this range image, the detailed scene characteristic is not effectively depicted due to the loss of the neighborhood relationship between the laser points. Bearing-angle images are another pictorial representation of 3-D laser data, which were proposed by Scaramuzza *et al.* to perform extrinsic calibration of a camera with a 3-D laser range finder [17].

Compared with range images, bearing-angle images show a superior performance in representing details of the scene, such as edges and corners in Fig. 2(c), which are very helpful for local feature extraction in our paper. A bearing angle is defined as the angle between the laser beam and the segment joining two consecutive measurement points [17]. Here, we consider the jth point in the ith 2-D scan $(x_{i,j}, y_{i,j}, z_{i,j})$ and the $(j+1)$th point in the $(i+1)$th 2-D scan $(x_{i+1,j+1}, y_{i+1,j+1}, z_{i+1,j+1})$ to be the adjacent measurement points. Our algorithm for building 45° bearing-angle images can be expressed as in (2), shown at the bottom of the next page, where $d\phi$ is the laser beam angular step in the direction of the trace and can be written as in (3), also shown at the bottom of the next page, where $a = x_{i,j}y_{i+1,j+1} - y_{i,j}x_{i+1,j+1}$, $b = y_{i,j}z_{i+1,j+1} - z_{i,j}y_{i+1,j+1}$, and $c = z_{i,j}x_{i+1,j+1} - x_{i,j}z_{i+1,j+1}$.

III. Extraction of Local SURF and Global Spatial Features

Our feature description of the indoor scene consists of two parts: 1) the local SURF features extracted from bearing-angle images; and 2) the global spatial features. By using these global

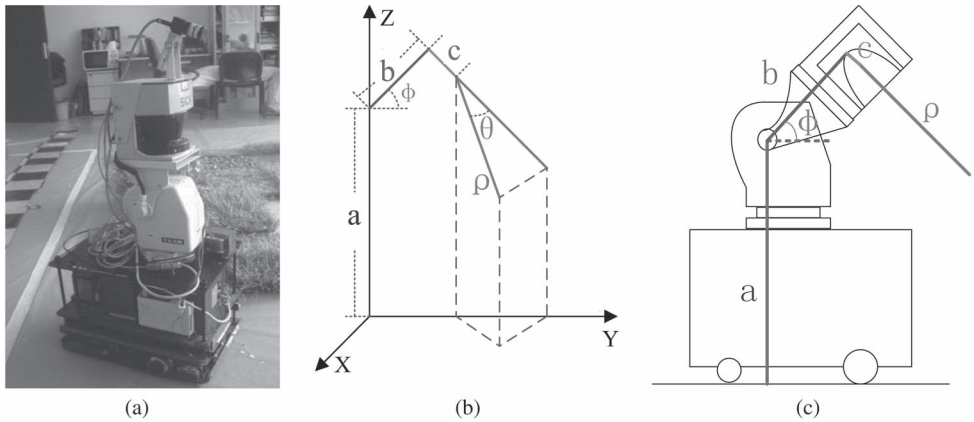


Fig. 1. Experiment equipment. (a) SmartROB2 mobile robot with a self-built 3-D laser scanner. (b) Corresponding polar coordinates. (c) Physical meaning of parameters in SmartROB2 and 3-D laser scanner.

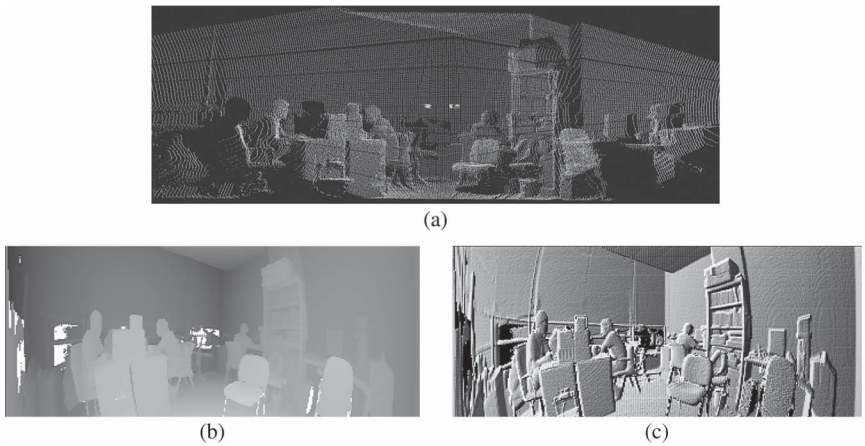


Fig. 2. Indoor scene is represented by raw 3-D laser point cloud, range image, and bearing-angle image, respectively. (a) Scene depicted by the raw 3-D laser scanning point cloud. (b) Range image of the same scene. (c) Bearing-angle image of the same scene.

and local features, large-scale indoor scenes can be described from coarse to fine.

A. Local SURF Features

The SIFT algorithm proposed by Lowe [18] in 1999 is a robust local image feature extraction method. It provides a robust descriptor and is able to adapt to scale changes, noise, rotations, and affine translations. In recent years, Bay and Tuytelaars [19] presented a fast and robust feature by using integral images for image convolutions and a fast Hessian detector, which is much faster than SIFT feature extraction.

SIFT descriptors extracted from the range image and the bearing-angle image are shown in Fig. 3(a) and (b), respectively. Correspondingly, SURF descriptors extracted from the same range image and the same bearing-angle image are shown in Fig. 3(c) and (d). Note that the SURF and SIFT features extracted from the bearing-angle image are more than those from the range image. A series of experiments have been conducted here to prove that the number of useful SURF or SIFT features from bearing-angle images is larger than that from range images. Although the number of SURF features is larger than that of SIFT features in the case of bearing-angle

$$\mathrm{BA}_{i,j} = \arccos \frac{\sqrt{x_{i,j}^2 + y_{i,j}^2 + z_{i,j}^2} - \sqrt{x_{i+1,j+1}^2 + y_{i+1,j+1}^2 + z_{i+1,j+1}^2} \cdot \cos d\phi}{\sqrt{(x_{i,j} - x_{i+1,j+1})^2 + (y_{i,j} - y_{i+1,j+1})^2 + (z_{i,j} - z_{i+1,j+1})^2}} \tag{2}$$

$$d\phi = \arcsin \frac{\sqrt{a^2 + b^2 + c^2}}{\sqrt{x_{i,j}^2 + y_{i,j}^2 + z_{i,j}^2} \cdot \sqrt{x_{i+1,j+1}^2 + y_{i+1,j+1}^2 + z_{i+1,j+1}^2}} \tag{3}$$

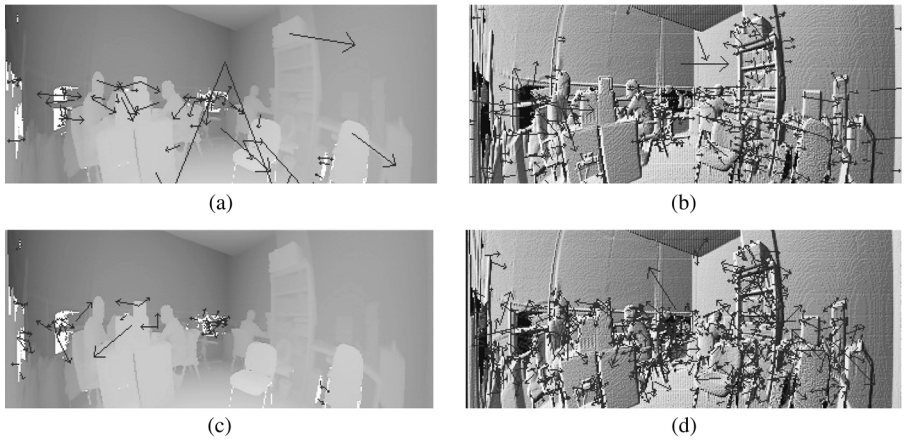


Fig. 3. Comparison of SIFT and SURF feature extraction. SIFT features extracted from (a) a range image and (b) a bearing-angle image. SURFs extracted from (c) the same range image and (d) the same bearing-angle image.

images, the number of useful SURF features are almost as many as that of useful SIFT features.

Although cluttered environments and people movements have significant effects on rotation and scale changing in bearing-angle images, SIFT and SURF can be both effectively used to obtain autonomous place recognition for a mobile robot in our paper. Since the use of SURF features needs less computation time than the use of SIFT features, a framework using multiple SURF features is deployed to identify different indoor places and is robust to partial occlusion. SURF features are stored in a k-d tree, and the best-bin-first strategy is used to reduce the computational cost for matching. In this process, the RANdom SAmple Consensus (RANSAC) [23] algorithm is used to remove mismatching pairs.

B. Global Spatial Features

If no additional scene information can be deployed in the process of recognition, we need to compare each SURF feature in the query scene with all the features in the database. The time for pair matching generally dramatically increases with the number of scenes in the database. Consequently, a fast detection algorithm based on 3-D spatial metric features is utilized to improve searching efficiency.

Since the main components of indoor scenes are a variety of rooms, the diverse scale parameters of these rooms can be used as global spatial features for selecting candidate scenes from a database. In this way, the search time mainly depends on the discrepancy of the global spatial features but not the scale of a database. Certainly, the worst case is that all scenes in a database have the same global spatial features. In this case, the search time is the same with the initial case without using global spatial features.

In this paper, there are about 150 2-D scans in each 3-D laser scan obtained by a mobile robot. Moreover, many obstacles, such as desks, chairs, and walking people will have an effect on the actual area of a 2-D scan. Since the maximal scanned area is usually obtained from the diagonal area of a scene, which can avoid these obstacles in the scene to a great extent, the maximal coverage area of all 2-D scans in a 3-D laser scan is selected as the global spatial feature to describe the scene. Therefore, we approximate a coverage area by using the following equations:

$$r_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2 + z_{ij}^2} \tag{4}$$

$$a_i = \frac{1}{2} \sin \alpha \left(\sum_{j=0}^{j=359} r_{ij} r_{i,j+1} \right) \tag{5}$$

where (x_{ij}, y_{ij}, z_{ij}) are the coordinates; r_{ij} is the range data of the jth laser point in the ith 2-D scan; α is the angular resolution, which is set to be 0.5° here; and a_i is the coverage area of the ith 2-D scan. As a result, the global spatial feature Area of a 3-D scan is defined as

$$\text{Area} = \max_i(a_i). \tag{6}$$

IV. Indoor Place Recognition Framework

A novel indoor place recognition framework is presented in Fig. 4. As shown, after 3-D laser scanning data of a query scene are obtained, the bearing-angle image is then created, and the global spatial feature is extracted from the original 3-D laser point cloud. By comparing the global spatial feature in the query scene with those in the database, priority rating of scenes in the database is accomplished. The candidate scenes are reordered and put in a queue. The scene at the head of the queue is considered as the optimal candidate scene.

To evaluate the similarity of these two scenes, local SURF features of the query scene are extracted from the bearing-angle image, which are used to match with those local SURF features in the candidate scene. The matching score can be calculated by using (9)–(11). If it is above the threshold of acceptance, the optimal candidate scene is regarded as the same as the query scene, and the place recognition is accomplished. Otherwise, if there are still unchecked candidate scenes, the suboptimal scene will be matched, or the query scene is considered to be a new one and added to the database. The detailed implementation of our indoor place recognition system is given in the following sections.

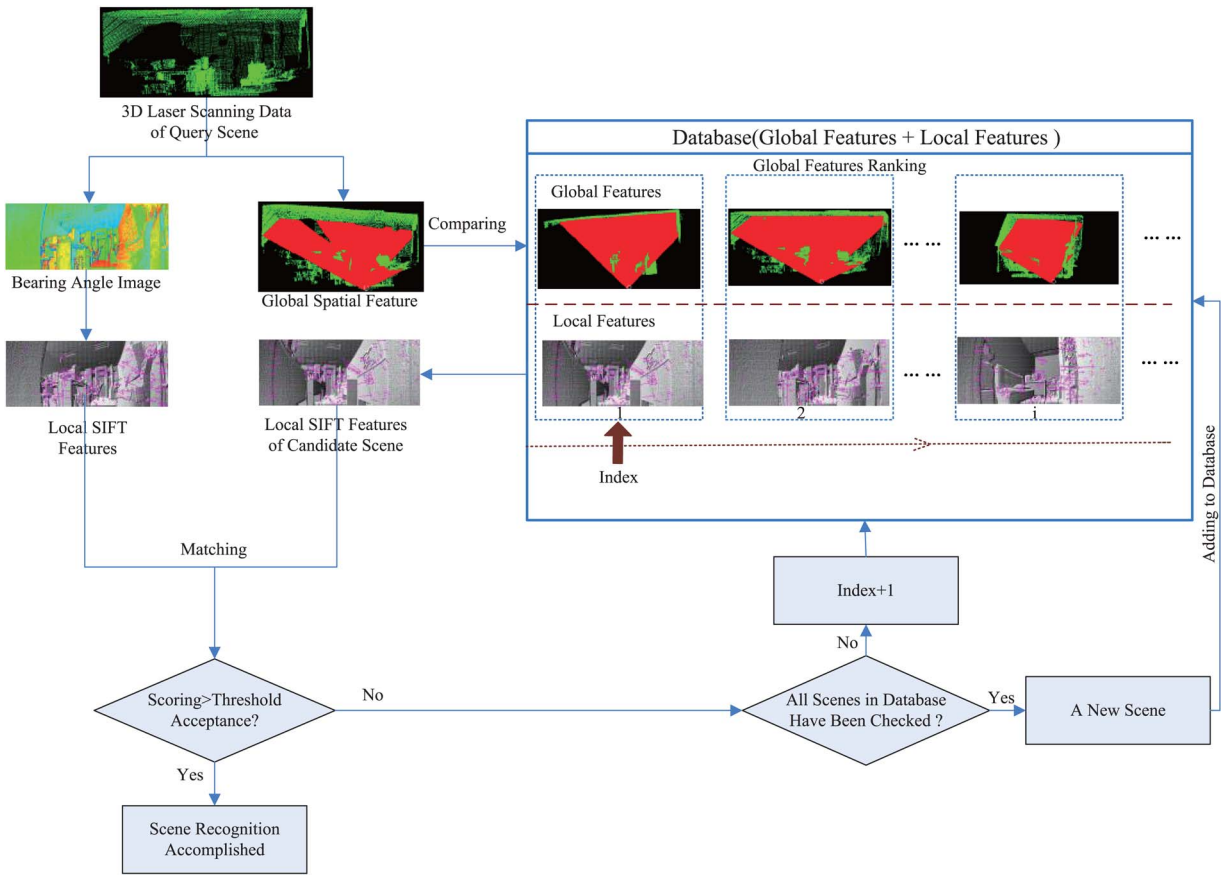


Fig. 4. Indoor place recognition framework.

A. Scene Priority Rating Based on Global Spatial Features

In order to speed up the recognition process, a scene priority-rating strategy based on global spatial features is utilized to avoid comparing the query scan with all the scenes in a database. Therefore, scenes in the database have to be ranked according to the ascending order of the differences of the global spatial features. Then, the scene in the front is selected as the optimal candidate one, and its SURF features can be used to judge whether it is the same scene as the query scene.

Taking the query scene in Fig. 5(a) for example, the ranking result based on global spatial features is shown in Fig. 5(b). The dark grey regions are the coverage areas of the 2-D scans used to calculate global spatial features. Note that the scenes with similar global features to the query scenes are all in the front and have higher priority to match. Experiments in our paper prove that less than three scenes in the database need to be matched with the query scene for recognizing the place by using this priority-rating strategy. Taking 30 scenes for example, it will cost about 40 ms to match the query scene with one scene in the database. Therefore, in the worst case, in which all scenes in the database should be compared, it will take almost 1.2 s for scene matching without priority-rating strategy. In contrast, no more than 120 ms will be taken with the help of priority-rating strategy. Overall, the search time can be significantly reduced to improve the real-time performance of the proposed method.

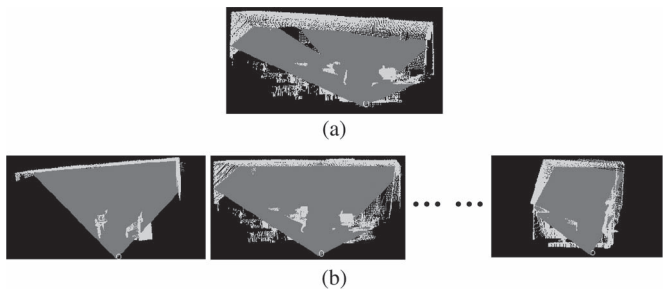


Fig. 5. Result of scene priority rating based on global spatial features. (a) Query scene. (b) Scenes in the database are ranked according to the ascending order of the differences of the global spatial features.

B. Calculating the Candidate Transformations

After the optimal candidate scene has been found, we can use the SURF feature correspondences between the optimal candidate scene and the query scene to calculate the candidate transformations. If at least three correct correspondences between two scans are given, we can calculate the corresponding 3-D transformation successfully [21]. In our paper, we use four correspondences to calculate a candidate transformation. There are mainly two methods widely used to calculate the transformation: unit quaternions [21] and singular-value decomposition (SVD) [22]. We adopt the SVD-based method in

this paper. The transformation is composed of a rotation matrix R and a translation vector T, i.e.,

$$R = \begin{bmatrix} r_0 & r_1 & r_2 \\ r_3 & r_4 & r_5 \\ r_6 & r_7 & r_8 \end{bmatrix} \qquad T = \begin{bmatrix} t_0 \\ t_1 \\ t_2 \end{bmatrix}. \tag{7}$$

For simplicity, $S = \{s_1, s_2, \cdots, s_n\}$ is defined as the matched feature list of the query scene and $D = \{d_1, d_2, \ldots, d_n\}$ as an optimal candidate scene. Now, we project the pixel associated with the SURF feature s_i into the 3-D laser point cloud, leading to the laser point p_i. Therefore, $P = \{p_1, p_2, \cdots, p_n\}$ can be defined as the laser point set corresponding to S, whereas $P' = \{p'_1, p'_2, \cdots, p'_n\}$ is the laser point set corresponding to D. We can define $C = \{c_1, c_2, \cdots, c_n\}$ as the 3-D laser point correspondence set, where $c_i(p_i, p'_i)$ is the ith correspondence.

To reject false matching pairs, we employ a variant of RANSAC [23] to detect an inlier set of maximal size. The process is described in Algorithm 1 in which parameter L should be calculated in advance. The probability of L consecutive failures is $p_{\text{fail}} = (1 - (p_g)^4)^L$, where p_g is the probability of a randomly selected data item being part of a good model, solving for $L = \log(p_{\text{fail}})/\log(1 - (p_g)^4)$ [23]. In our paper, we set $p_g = 0.8$ and $p_{\text{fail}} = 0.0001$, and solve for $L = 17.48$. If we set $L = 18$, the probability of 18 consecutive failures is below 0.0001, which meets our need.

Algorithm 1. A variant of RANSAC algorithm to detect an inlier set of maximal size

Generate correspondence list C_{list} and C_{temp}
Integer k
for $k = 1$ to L **do**
  clear C_{temp}
  select a quadruple of correspondences from set C randomly
  calculate (R, T) based on SVD method
  register P into P'' :$P'' = PR + T$
  for $i = 1$ to n **do**
    if dis$(p'_i, p''_i) \leq 0.15$
      add $c(p_i, p'_i)$ to C_{temp}
    end if
  end for
  if $|C_{\text{temp}}| > |C_{\text{list}}|$
  $C_{\text{list}} = C_{\text{temp}}$
  end if
end for
Get the inlier set $C_{\text{list}} = \{c_1, c_2, \cdots, c_{n'}\}$
$p_i \in P$; $p'_i \in P'$; $p''_i \in P''$.

Quadruples of correspondences are then selected from the inlier set C_{list} based on the Good Sample Consensus (GOODSAC) algorithm [24] to improve the computation accuracy of the transformation matrix when the distances of the quadruples of correspondences are too close.

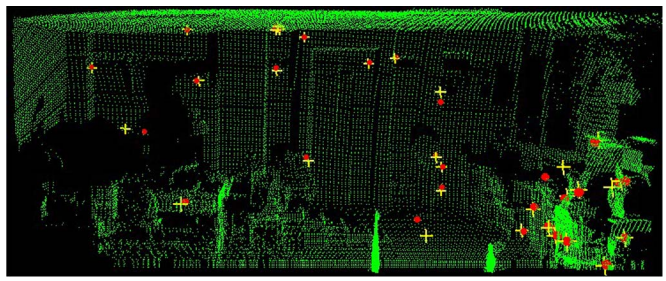


Fig. 6. Matching pairs between the query scene and the one in the database.

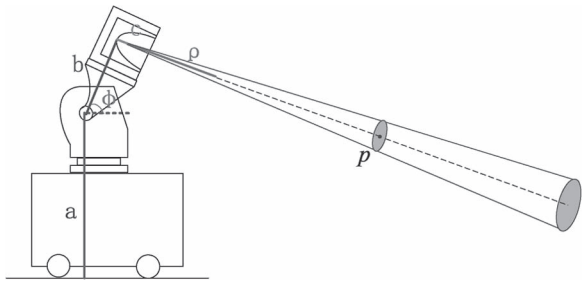


Fig. 7. Cone neighborhood of a laser point.

Thereby, we obtain the quadruple list $C_{\text{Quadruple}} = \{(c_0, c_1, c_2, c_3), (c_0, c_1, c_2, c_4), (c_0, c_1, c_2, c_5), \cdots\}$. (c_i, c_j, c_m, c_n) in $C_{\text{Quadruple}}$ is used to calculate a candidate transformation (R_{ijmn}, T_{ijmn}) using the SVD-based method, and then, the transformation list $RT = \{(R_0, T_0), (R_1, T_1), (R_2, T_2) \cdots\}$ is obtained. It should be pointed out that the number of possible combinations exponentially grows with the number of correspondences. In practical applications, we only select at most 50 transformations from RT to reduce the computational cost of scoring. As correct transformations are mutually consistent, we can use K-means clustering [25] to obtain the rational ones. RT is clustered into two categories, and then, the bigger one will repeat the process until the number of transformations is below 50. The remains of RT are regarded as the best candidate transformations.

By using the K-means algorithm, we can obtain the best candidate transformations. Matching pairs between two bearing-angle images are mapped to raw 3-D point cloud of a scene in the database, as depicted in Fig. 6. White crosses represent SURF features of the scene in the database, whereas the dark grey points are matched SURF features in the query scene.

C. Scoring of the Candidate Transformations

Each candidate transformation is calculated only by four feature correspondences. The validity of the candidate transformation is checked by all the other feature correspondences. A set of 3-D points P belonging to the query scene can be transformed to a new set $P'' = \{p''_1, p''_2, \cdots, p''_n\}$ belonging to the optimal candidate scene by using the candidate transformation; thus, we can compare P' and P'' in the same coordinate. To avoid the influences of small errors between P' and P'', the set of cone neighborhood is adopted, so that the matching pairs can be correctly estimated, particularly when one point occurs

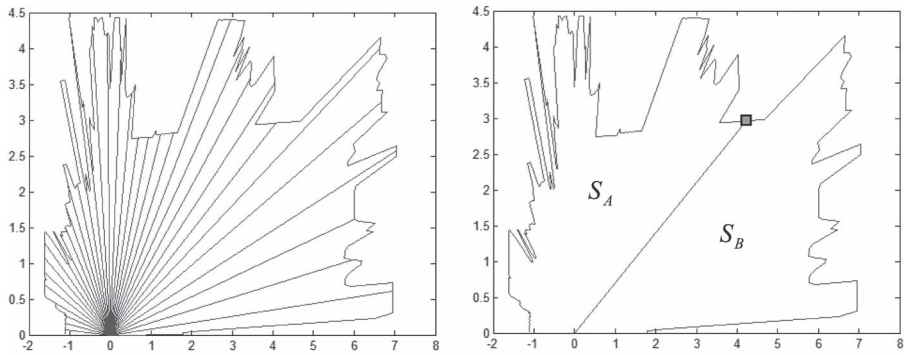


Fig. 8. Experimental results show how the guideline of the 3-D laser scanning direction is decided while a robot enters into a room. (a) The irregular horizontal area is divided into a series of small regions according to equal scanning angle intervals. (b) The whole area is segmented into two big regions with equivalent area approximately.

at the edges of an object and the other falls far away on the background.

For a laser point p, its cone neighborhood set is defined as

$$N = \{p_i | \quad |\text{Angle}(p_i, p)| \leq \Theta\} \tag{8}$$

where $\text{Angle}(p_i, p)$ is the angle between the two scanning lines in which points p_i and p lie on, and Θ is the maximum allowed error (we used 3° in our paper). As shown in Fig. 7, the dashed line indicates the laser scanning line in which the laser point p lies on, and the cone region enclosing this point is its cone neighborhood area.

Next, we will calculate a score for each pair (p_i'', p_i') in P'' and P'. If p_i'' is not in the cone neighborhood area of p_i', the score should be 0, since this pair is considered as a mismatch. Otherwise, this pair is valid, and the score will be between 0 and 1, which is relative to $\text{Angle}(p_i'', p_i')$ and the distance from the optical center of laser range finder to point p_i'. Therefore, the scoring function for a single pair is defined in (9), shown at the bottom of the page, where $D_{p_i''}$ and $D_{p_i'}$ are the distance from the optical center of the laser scanner to the laser points p_i'' and p_i', respectively, $D_{\max}$ is the maximal distance from the optical center of the laser scanner to a laser point in the scene, and $1/[1 + (\min\{D_{p_i''}, D_{p_i'}\}/D_{\max})^{2N}]$ is a weighting function for eliminating the inaccuracy introduced by the range (the function's selection is based on the Butterworth low-pass filter and $N = 3$ in our paper).

After the scores of all the pairs have been calculated, the score of the candidate transformation can be obtained as follows:

$$S = \frac{\sum_{i=1}^{n} s_i}{n} \tag{10}$$

where n is the number of the valid pairs.

To avoid the case where several good valid pairs in two different scenes result in a high score, we modify (10) as follows:

$$S = \begin{cases} \dfrac{\sum_{i=1}^{n} s_i}{n} & n \geq 17 \\ \dfrac{\left(\sum_{i=1}^{n} s_i\right) * e^{\frac{(n-17)^2}{3\times 3}}}{n} & n < 17. \end{cases} \tag{11}$$

If the number of valid pairs is below 17, a Gaussian penalty is given to S. If the highest score is above a threshold, the scenes according to the two sets P' and P'' are believed to be the same scene.

V. Experiments and Analysis

A. Experiment Setup

The proposed indoor place recognition approach has been implemented on a SmartROB2 mobile robot and tested in a series of indoor office scenes. The testing environment is on the sixth floor of the Chuangxinyuan building at Dalian University of Technology, which has more than 30 offices and laboratories for the evaluation of the proposed algorithm.

In practical applications, when mobile robots fail to localize themselves, they should have the ability to recover from failure rapidly. Therefore, place recognition is a good choice to help a mobile robot know where it is. Due to the scene similarity and dynamic disturbances in indoor corridor environments, it is better to let a mobile robot enter a room to accomplish place recognition. The door detection algorithm used in this paper can be found in our previous work [20].

As a robot goes through the door and enters the room, it will estimate the free space in the room only using its laser scanning data in the horizontal plane. As shown in Fig. 8(a), this irregular horizontal area is divided into a series of small regions according to equal angle intervals of the laser scanning

$$s_i = \begin{cases} 0 & \text{if } p_i'' \text{ is not in the neighborhood area of } p_i' \\ \left(1 - \frac{|\text{Angle}(p_i'', p_i')|}{\Theta}\right) \cdot \frac{1}{1+(\min\{D_{p_i''}, D_{p_i'}\}/D_{\max})^{2N}} & \text{if } p_i'' \text{ is in the neighborhood area of } p_i' \end{cases} \tag{9}$$

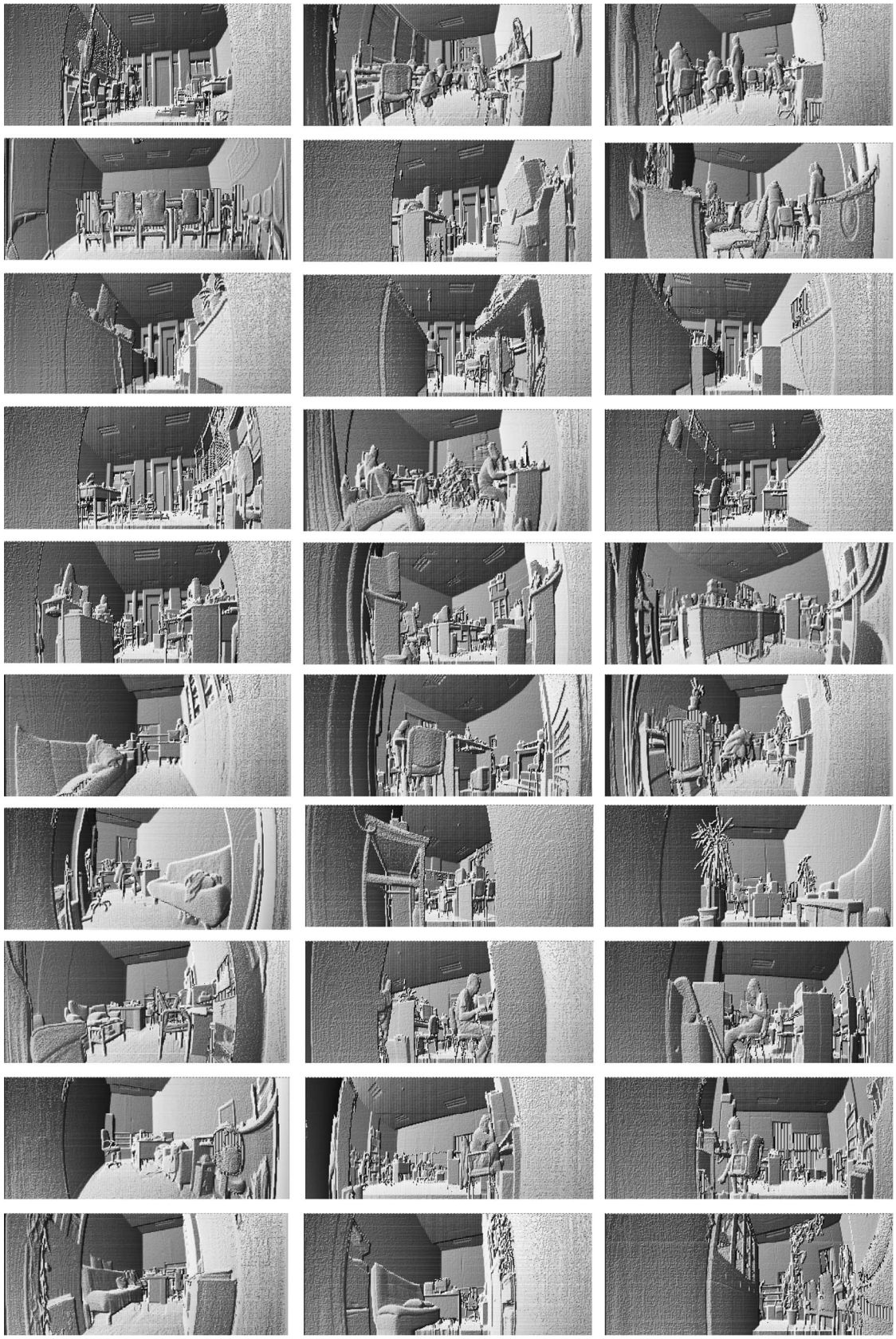


Fig. 9. Bearing-angle images of 30 scenes in the database.

(divided at intervals of 5° in our paper). Then, the robot can segment the whole area into two big regions with equivalent area approximately. As shown in Fig. 8(b), $S_A \approx S_B$ and the cut line will be used as the guideline of the 3-D laser scanning direction while autonomous place recognition is performed in dynamic indoor environments. With the help of the guideline,

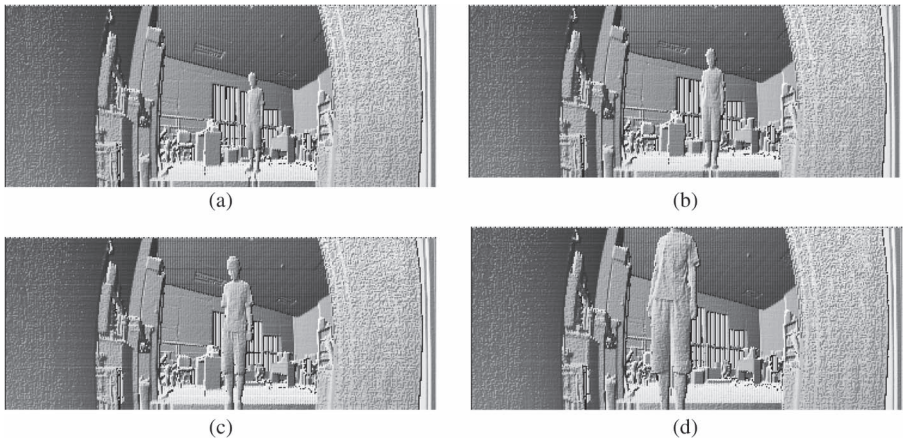


Fig. 10. Four testing scenes obtained at the same office, where a person stood at (a) 4, (b) 3, (c) 2.5, and (d) 1.5 m from the mobile robot.

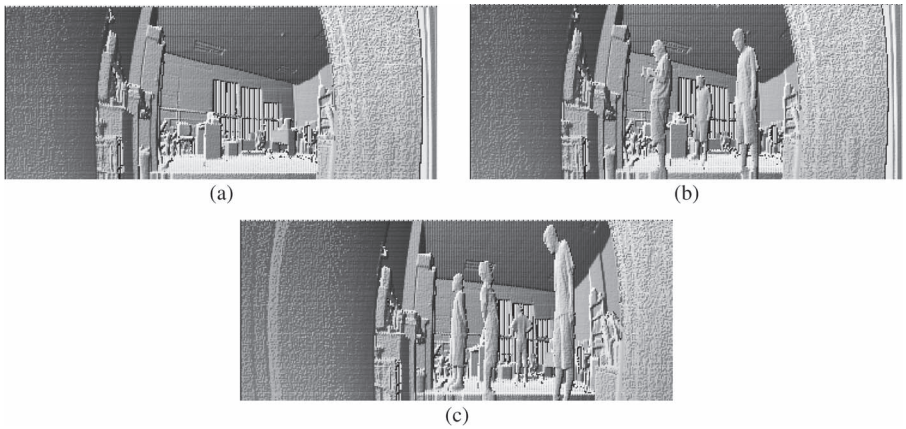


Fig. 11. Occlusion of multiple persons tests. (a) No people in the scene. (b) Three and (c) four people in the scene.

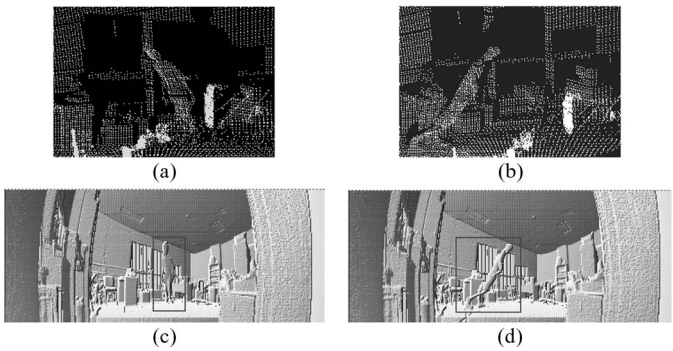


Fig. 12. Dynamic disturbances tests. One person is walking toward (a) left and (b) right in scans of the indoor scene. The corresponding bearing-angle images are given in (c) and (d).

a mobile robot could keep the similar orientation after entering a room, so that the similar field of view for 3-D laser scanning will be obtained in place recognition.

B. Database Construction and Confusion Matrix

The database is composed of local SURF features and global spatial features extracted from 30 indoor scenes. Bearing-angle images of these 30 scenes are shown in Fig. 9, and an average of about 400 SURF features can be extracted from every scene's bearing-angle image in the database. To evaluate the divergence

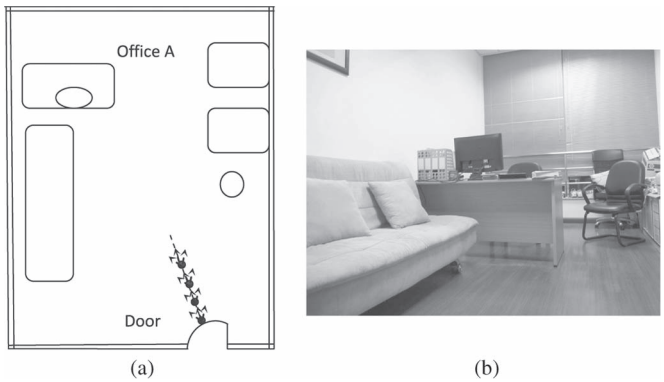


Fig. 13. Office A for testing. (a) Layout of office A. (b) Corresponding image of this office.

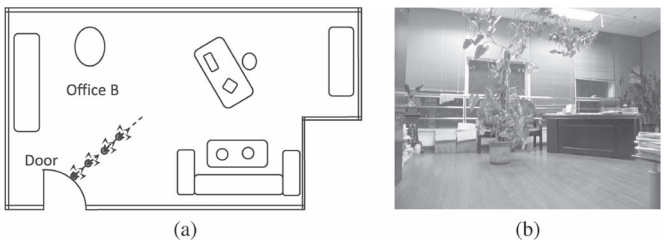


Fig. 14. Office B for testing. (a) Layout of office B. (b) Corresponding image of this office.

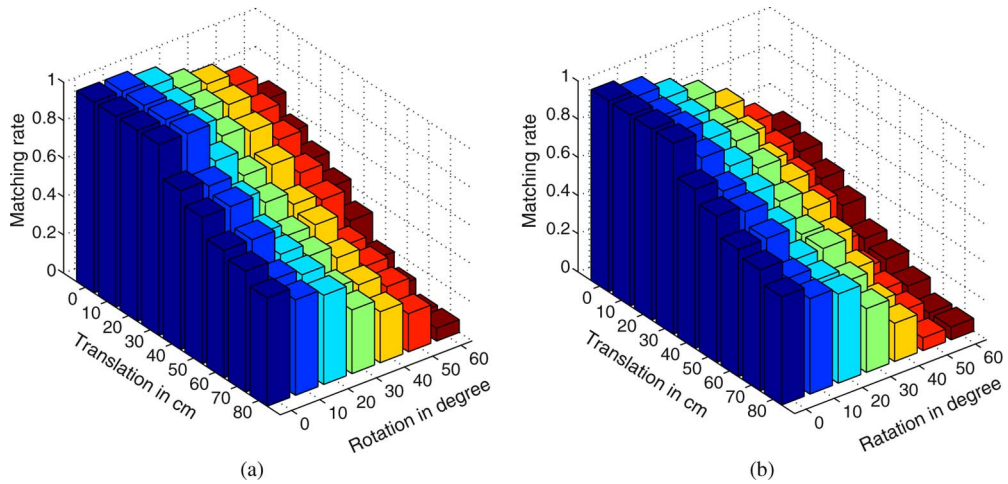


Fig. 15. Translation and rotation limit testing in offices A and B. Test results in offices (a) A and (b) B.

between scenes in the database, we calculated the confusion matrix, and each scene is matched against the others. The result shows that, when matching the database against itself, we had a 100% detection rate with zero false positives.

C. Robustness

In general, there are mainly two kinds of disturbances in our experiments: people's presence and movements. As shown in Fig. 10, four testing scenes were obtained at the same office, where a person stood at different distances to the robot. The occlusion area increases when the distance between the person and the robot decreases, i.e.,(4, 3, 2.5, and 1.5 m), which made place recognition difficult. Now, we match the aforementioned scenes with the ones in the database, and their matching rates with the same scene are 92.24%, 91.85%, 93.23%, and 91.89%, respectively (the acceptance threshold used in our work is 50%). It was found from the experiments that the reliable distance for people's occlusion is no less than 1.5 m.

Another experiment is to test the occlusion of multiple persons, which causes a significant variance among bearing-angle images. Compared with the original scene [see Fig. 11(a)] in the database, three and four people appeared in the testing scenes [see Fig. 11(b) and (c)]. The matching rate are 91.12% and 72.87% for Fig. 11(b) and (c), respectively. The recognition result shows that this method has good performance under the condition of foreground changes.

There is another common situation that people are walking in the scene while the mobile robot is taking a scan. Laser scanning data in Fig. 12(a) show a person walking toward the robot, whereas Fig. 12(b) depicts a person walking across the room. Fig. 12(c) and (d) are the corresponding bearing-angle images, and the black rectangular boxes in Fig. 12(c) and (d) show the dynamic disturbances caused by walking human targets in different cases. The matching results are 90.81% and 78.83%, which demonstrate the robustness of the recognition approach to dynamic disturbance.

Possible changes of laser-scanner observing locations may cause significant changes in local and global features, and make the place recognition a big challenge. To test whether SURF features in the same spatial area could be stably recognized with some tolerance against view angle and distance between different observing locations, a series of quantitative tests is given in two typical offices shown in Figs. 13 and 14. There are notable differences between the layouts of these two offices. Both the translation in view direction and the rotation on the spot are simultaneously arranged in these tests. The testing results are given in Fig. 15. The rotation and translation tolerances in office A are 0.9 m and 30°, whereas the rotation and translation tolerances in office B are 0.8 m and 20°. According to these results of the rotation and translation limit testing, we can define the safety threshold of the translation and rotation tolerances as 0.8 m and 20°, which is sufficient for our implementation. Fig. 16 shows the safe working region for mobile robots. The appropriate translation and rotation can be chosen below the curve in this figure.

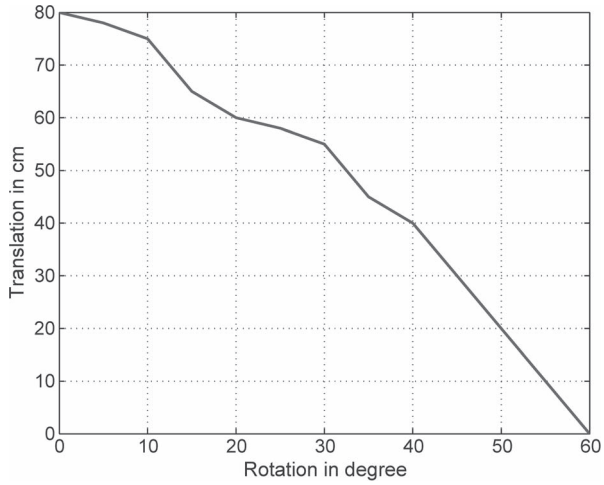


Fig. 16. Safe working region for mobile robots.

Figs. 17 and 18 show more experimental results to further prove that the proposed place recognition system can cope with disturbances effectively. As shown in Fig. 17, four disturbances are simultaneously considered in order to make these tests more realistic, namely, the translation in view directions, the rotation on the spot, static disturbances (e.g., people standing in the foreground of the testing scene), and the dynamic disturbance

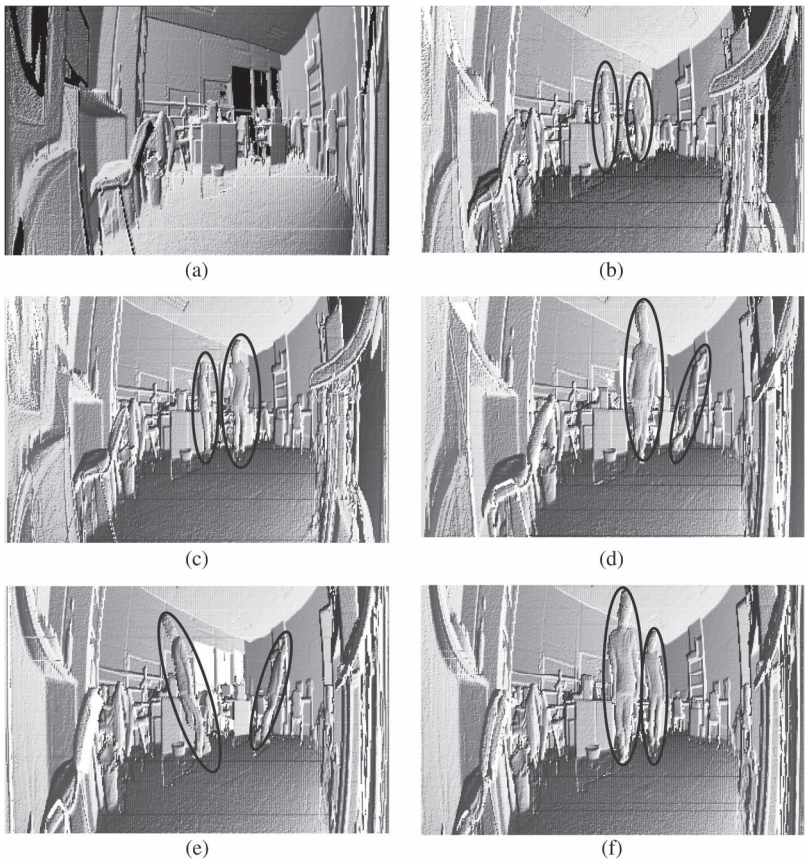


Fig. 17. Some more realistic disturbances tests, including the translation in view direction, the rotation on the spot, static disturbances, and the dynamic disturbance simultaneously. (a) Original scene. (b) Two persons standing in the foreground. (c) One person standing and one person walking in the foreground. (d) One person walking forward and one person walking across in the foreground. (e) Two persons walking across in the foreground simultaneously. (f) Two persons walking forward in the foreground simultaneously.

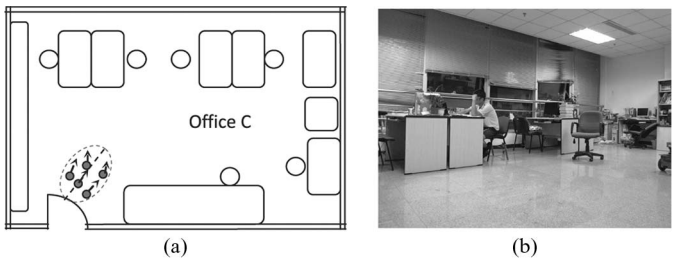


Fig. 18. Office C for testing. (a) Layout of office C. (b) Corresponding image of this office.

TABLE I
PLACE RECOGNITION EXPERIMENT RESULTS

Scene ID	a	b	c	d	e	f
Translation (cm)	$(0, 0)$	$(10, -20)$	$(-10, -20)$	$(-20, 20)$	$(10, 20)$	$(20, 20)$
Rotation (°)	0	20	20	30	20	30
Matching score	1	0.86	0.91	0.87	0.87	0.67

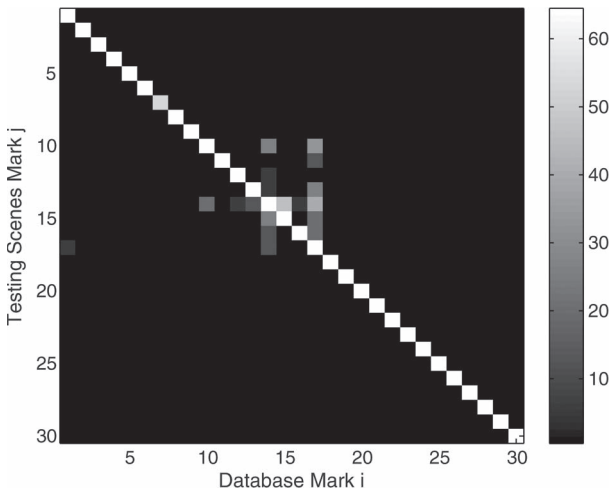


Fig. 19. Place recognition result of 30 scenes in indoor environment.

(e.g., people continuously walking in the foreground of the testing scene). As shown in Fig. 18(a), the black dashed ellipse illustrates the effective region for the robot to obtain laser scanning data, whereas the dark grey points with arrows illustrate the spots and the view directions of the robot.

The corresponding bearing-angle images are given in Fig. 17, and black ellipses in these images illustrate the positions of different people. As shown in Fig. 17(b)–(f), two persons are standing and walking across or forward simultaneously in different situations in the foreground. The place recognition experiment results of these cases are given in Table I. As shown, the matching scores are between 0.67 and 1.0. These experimental results show that the place recognition system in our robot can deal with dynamic disturbances in the foreground of the testing scene effectively.

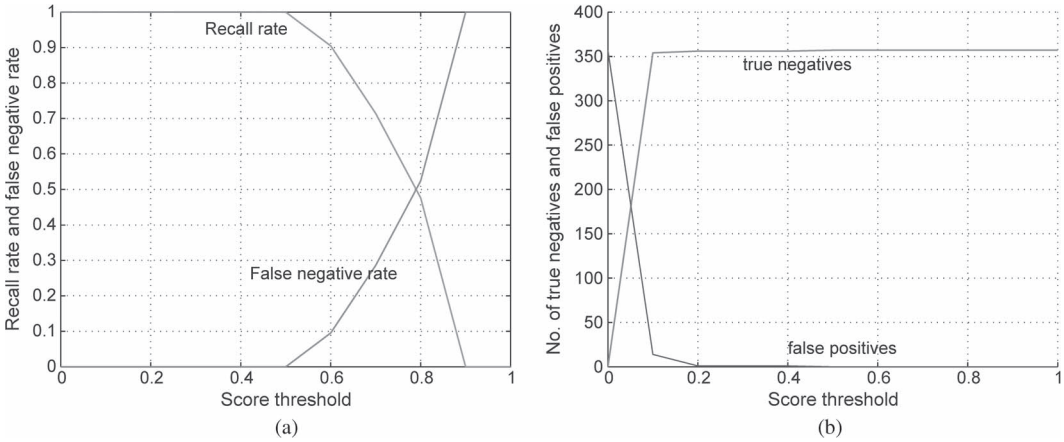


Fig. 20. Indoor place recognition result. (a) Percentage of true positive and false negative of 30 scenes to be recognized. (b) Number of true negatives and false positives of 30 scenes.

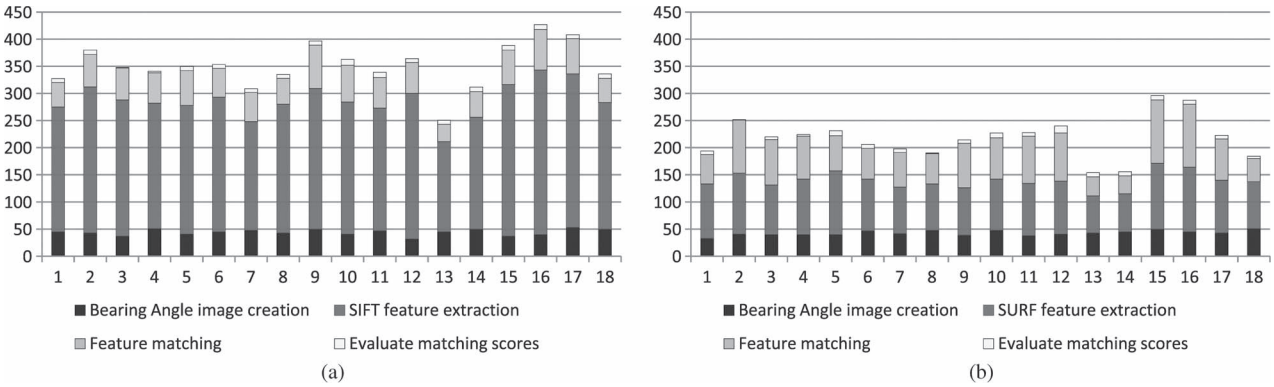


Fig. 21. Time cost for matching a scan against 18 scenes selected from the database randomly. Time cost of place recognition based on (a) SIFTs and (b) SURFs.

D. Recognition Results

Thirty scenes are obtained to match with those in the database, and the result is shown in Fig. 19. It is clear that the matching rate of the same scene is much higher than the others. If we set the acceptance threshold to be 0.5, these scenes could be correctly recognized. Please note that the scenes in Fig. 19 are acquired at different positions in the same room, and they are recognized as the same scene in the experiments.

Fig. 20 is the statistical analysis result of Fig. 19. There is no false positive when the acceptance threshold is larger than 0.47 [see Fig. 20(b)]. In fact, there is one false positive if the acceptance threshold is between 0.2 and 0.47 in our experiments. Fig. 20(a) indicates that, when the acceptance threshold is below 0.6, the recall rate is above 90%. Thus, the acceptance threshold should be set between 0.47 and 0.6, and here, 0.5 was chosen in our experiments. Note that there is no false positive while the score is above 0.47.

E. Timing

The runtime analysis is carried out on a Core2 Duo 1.83-GHz central processing unit, and the result is shown in Fig. 21. Both SIFT and SURF features are tested for comparing the time cost based on 18 scenes that are randomly selected from 30 scenes. We need no more than 55 ms to calculate the bearing-angle image. In the process of local features extraction, about 170−300 ms is used to extract SIFT features for each scan, whereas only about 70−120 ms is used to extract SURF features for each scan. The k-d tree is introduced to speed up the searching for matching pairs, and it costs no more than 130 ms to build a k-d tree and search out the matching pairs.

In our paper, the time cost in querying the database is not linear with the number of scenes in the database because of using the scene priority-rating strategy. For any query scene, it only needs to match with three scenes in a database on average, which takes about 0.23 s on average to accomplish autonomous place recognition. This is acceptable compared with the time of taking a scan.

VI. Conclusion and Future Work

This paper has focused on how to accomplish indoor scene measurement and autonomous indoor place recognition by a mobile robot based on 3-D laser scanning data. In order to obtain scale-invariant features from bearing-angle images, the SURF algorithm has been used for feature extraction and matching, which could provide local feature descriptors more robustly, even in dynamic indoor environments. Moreover, global spatial features are also used to judge the scene priority rating, which can reduce the computational cost effectively. By using both the local SURF features and the global spatial features, a novel place recognition framework with a diverse database is proposed for handling autonomous indoor place recognition. Experimental results have shown the validity and the robustness of the proposed method.

Our future work will further test this place recognition approach in large-scale unstructured environments and improve its practicability and performance. Furthermore, other feature descriptors that could result in better accuracy and lower computational cost in the real-world applications will be investigated.

REFERENCES

[1] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *Proc. IEEE Conf. CVPR*, Miami, FL, 2009, pp. 413–420.

[2] M. Cummins and P. Newman, "Highly scalable appearance-only SLAM FAB-MAP 2.0," in *Proc. RSS*, 2009, pp. 1–8.

[3] M. Liu, D. Scaramuzza, C. Pradalier, R. Siegwart, and Q. Chen, "Scene recognition with omnidirectional vision for topological map using lightweight adaptive descriptors," in *Proc. IEEE/RSJ Int. Conf. IROS*, St Louis, MO, Oct. 2009, pp. 116–121.

[4] A. C. Murillo and J. Kosecka, "Experiments in place recognition using gist panoramas," in *Proc. IEEE Workshop Omnidirect. Vis., Camera Netw. Non-Classical Cameras, ICCV*, Kyoto, Japan, 2009, pp. 2196–2203.

[5] C. Cadena, D. Gálvez-López, F. Ramos, J. D. Tardós, and J. Neira, "Robust place recognition with stereo cameras," in *Proc. IEEE/RSJ Int. Conf. IROS*, Taipei, Taiwan, Oct. 2010, pp. 5182–5189.

[6] S. Hussmann and T. Liepert, "Three-dimensional TOF robot vision system," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 1, pp. 141–146, Jan. 2009.

[7] D. A. Grejner-Brzezinska, C. K. Toth, H. Sun, X. Wang, and C. Rizos, "A robust solution to high-accuracy geolocation: Quadruple integration of GPS, IMU, pseudolite, and terrestrial laser scanning," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 11, pp. 3694–3708, Nov. 2011.

[8] J. Ryde and H. Hu, "3D mapping with multi-resolution occupied voxel lists," *Auton. Robots*, vol. 28, no. 2, pp. 169–185, Feb. 2010.

[9] J. L. L. Galilea, J.-M. Lavest, C. A. L. Vazquez, A. G. Vicente, and I. B. Munoz, "Calibration of a high-accuracy 3-D coordinate measurement sensor based on laser beam and CMOS camera," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 3341–3346, Sep. 2009.

[10] L. Gatet and H. Tap-Beteille, "Measurement range increase of a phase-shift laser range finder using a CMOS analog neural network," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 6, pp. 1911–1918, Jun. 2009.

[11] M. Magnusson, H. Andreasson, A. Nuchter, and A. J. Lilienthal, "Appearance-based loop detection from 3D laser data using the normal distributions transform," in *Proc. IEEE ICRA*, Kobe, Japan, May 2009, pp. 23–28.

[12] B. Steder, G. Grisetti, and W. Burgard, "Robust place recognition for 3D range data based on point features," in *Proc. IEEE ICRA*, Anchorage, AK, 2010, pp. 1400–1405.

[13] B. Steder, G. Grisetti, M. Van Loock, and W. Burgard, "Robust online model-based object detection from range images," in *Proc. IEEE/RSJ Int. Conf. IROS*, St. Louis, MO, Oct. 2009, pp. 4739–4744.

[14] A. E. Johnson and M. Hebert, "Surface registration by matching oriented points," in *Proc. Int. Conf. Recent Adv. 3-D Digit. Imag. Model.*, Ottawa, ON, Canada, 1997, p. 121.

[15] Y. Zhuang, Y. Li, and W. Wang, "Robust indoor scene recognition based on 3D laser scanning and bearing angle image," in *Proc. IEEE ICRA*, Shanghai, China, 2011, pp. 4042–4047.

[16] C. Taylor and A. Cowley, "Fast scene analysis using image and range data," in *Proc. IEEE ICRA*, Shanghai, China, May 2011, pp. 3562–3567.

[17] D. Scaramuzza, A. Harati, and R. Siegwart, "Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes," in *Proc. Int. Conf. IROS*, San Diego, CA, Oct. 29–Nov, 2, 2007, pp. 4164 –4169.

[18] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. ICCV*, Kerkyra, Greece, Sep. 1999, pp. 1150–1157.

[19] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.*, Graz, Austria, May 2006, pp. 404–417.

[20] Y. Zhuang, X. Lu, Y. Li, and W. Wang, "Mobile robot indoor scene cognition using 3D laser scanning," *Acta Autom. Sin.*, vol. 37, no. 10, pp. 1232–1240, 2011.

[21] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Amer. A*, vol. 4, no. 4, pp. 629–642, Apr. 1987.

[22] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least square fitting of two 3-D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987.

[23] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[24] E. Michaelsen, W. von Hansen, M. Kirchhof, J. Meidow, and U. Stilla, "Estimating the essential matrix: GOODSAC versus RANSAC," in *Proc. Symp. PCV*, 2006, pp. 1–6.

[25] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, 1967, vol. 1, pp. 281–297, Univ. California Press, Berkeley, CA.

Yan Zhuang received the B.S. and M.S. degrees in control theory and engineering from Northeastern University, Shenyang, China, in 1997 and 2000, respectively, and the Ph.D. degree in control theory and engineering from Dalian University of Technology, Dalian, China, in 2004.

He was a Lecturer in 2005 and an Associate Professor in 2007 with Dalian University of Technology, where he is currently a Professor with the School of Control Science and Engineering. His research interests include mobile-robot localization, 3-D mapping, and place recognition.

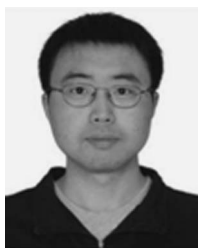
Nan Jiang received the B.S. degree in control theory and engineering in 2010 from Dalian University of Technology, Dalian, China, where he is currently working toward the M.S. degree in control theory and engineering.

His research interests include mobile-robot scene recognition, human detection, and object tracking.

Huosheng Hu (M'94–SM'01) received the M.Sc. degree in industrial automation from the Central South University, Changsha, China, in 1982 and the Ph.D. degree in robotics from the University of Oxford, Oxford, U.K., in 1993.

Currently, he is a Professor with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K., leading the Human-Centred Robotics Group. His research interests include behavior-based robotics, human-robot interaction, service robots, embedded systems, data fusion, learning algorithms, mechatronics, and pervasive computing. He has authored around 360 papers in journals, books, and conferences in these areas.

Prof. Hu was a recipient of a number of best paper awards. He is a founding member of IEEE Robotics and Automation Society Technical committee on Networked Robots, a Fellow of the Institution of Engineering and Technology and the Institute of Measurement and Control, and a Senior Member of the Association for Computing Machinery. He has been a Program Chair or a member of Advisory/Organizing Committee for many international conferences such as the IEEE International Conference on Robotics and Automation, Intelligent Robots and Systems, International Conference on Mechatronics and Automation, International Conference on Robotics and Biomimetics (ROBIO), International Conference on Automation and Information, International Conference on Automation and Logistics; and the International Association of Science and Technology for Development Robotics and Applications, Control and Applications, and Computational Intelligence Conferences. He currently serves as the Editor-in-Chief for the International Journal of Automation and Computing, the Executive Editor for the International Journal of Mechatronics and Automation, and the Editor-in-Chief for the Robotics Journal.



Fei Yan received the B.S. degree in control theory and engineering and the Ph.D. degree from Dalian University of Technology, Dalian, China, in 2004 and 2011, respectively.

Currently, he is a Postdoctoral Research Associate with the Robotics Laboratory, The City College, City University of New York, New York. His research interest covers 3-D environment modeling and path planning.