Robust Localization System Fusing Vision and Lidar Under Severe Occlusion

YONGLIANG SHI[ID], **WEIMIN ZHANG**[ID], **FANGXING LI**[ID], **AND QIANG HUANG**[ID]

School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China
Key Laboratory of Biomimetic Robots and Systems, Ministry of Education, Beijing Institute of Technology, Beijing 100081, China
Beijing Advanced Innovation Center for Intelligent Robots and Systems, Beijing 100081, China

Corresponding author: Weimin Zhang (zhwm@bit.edu.cn)

ABSTRACT Localization is one of the most fundamental problems for mobile robot. Aiming at the phenomenon that robot is prone to be lost in navigation under severe occlusion, a robust localization system combining vision and lidar is proposed in this paper. The system is split into off-line stage and online stage. In the off-line stage, this paper introduces a method of actively detecting and recording visual landmarks, and an off-line visual bag-of-words is generated from the recorded landmarks training. In the online stage, the prediction and update phase of Adaptive Monte Carlo Localization (AMCL) are improved respectively to enhance the performance of localization. The prediction phase generates the proposal distribution according to the prior information obtained through retrieving visual landmarks, and the newly proposed measurement model that selects reliable beams of lidar as the observation is to update the prediction. Experiments is carried out under strict conditions, that is 60% of the lidar is occluded, 1/12 of the beams are regarded as observation, and only 300 particles were adopted at most, it is shown that, no matter in the global localization or pose tracking, the localization system proposed in this paper performs much better than the state of art localization algorithm AMCL.

INDEX TERMS Robot localization, severe occlusion, proposal distribution, measurement model.

I. INTRODUCTION

Localization against a global map, including global localization and pose tracking, is a prerequisite for any robotics task where a robot must know where it is. In the past, a variety of approaches in fusion of multi-sensors for mobile robot localization have been developed [1]–[5]. The Markov localization method requires excessive computational overhead and a priority commitment to the size and resolution of the state space [6], which is difficult to meet the real-time requirements; The localization efficiency based on Kalman filter (KF) is high, but the Kalman filter is based on the assumption of linear Gaussian model, and it is difficult to describe the motion model and observation model accurately with linear model; On this basis, extended Kalman filter (EKF) and unscented Kalman filter algorithm are introduced [7], [8], these two methods solve the state transfer by linearizing the non-linear model of the original system near the operating point by first-order Jacobian approximation [9]. These two methods perform well in the pose estimation of the robot under certain

conditions, but will get worse estimation results in the position with poor linear conditions. Consequently, particle filter is introduced to solve this problem because it can approximate any complex multimode probability distribution [10]. In the prediction step, one sensor, for an example encoder, is used to transfer the prior state of motion, and the other sensor is used as observation model to update the prior state to realize correction step. On the basis of particle filter, Thrun *et al.* [6] introduced AMCL algorithm, Kullback-leibler Divergence (KLD) sampling makes the localization more efficient [11], so the AMCL is the most widely used approach in robot navigation. As the state-of-art localization method base on lidar, the AMCL has obvious advantages in efficiency, stability and accuracy, but it does not perform well in global localization. Meanwhile, the it is easy to introduce erroneous observations under severe occlusion conditions leading to failure of localization.

In this paper, an integrated localization system is proposed to solve the problem that lidar-based method is prone to failure in localization under severe occlusion. In order to automatically obtain valid landmarks, Rao-Blackwellized Particle Filters (RBPF) [12] is combined with orb-based

---

The associate editor coordinating the review of this manuscript and approving it for publication was Chenguang Yang[ID].

visual tracking methods is firstly proposed to actively detect and record valid landmarks, and the recorded landmarks are trained through kmeans++ to generate an off-line visual database of bag-of-words (DBoW) [13], [14]. In prediction phase of global localization, the visual DBoW is used to search first to obtain the most similar landmark used as reliable prior to generate proposal distribution, while the encoder is to provide prior in prediction phase of pose tracking. Moreover, a newly proposed lidar-based measurement model is used in correction phase of both global localization and pose tracking, which select the mostly reliable beams as observation to update state of particles. The advantage of this model is that it can avoid introducing both dynamic and static obstacle information into observation model to the greatest extent. Due to the improvement of the prediction step and correction step, the performance of localization for robot under severe occlusion is greatly promoted. The main contributions are as follows.

1. An active landmark detection and recording method is proposed, and the concept of matching rate is introduced as an evaluation criterion, which solves the problem of artificial setting of the landmarks.
2. DBoW based visual retrieval method provides a reliable prior for global localization to generate proposal distribution instead of randomly selecting an initial position in prediction step of global localization;
3. A new lidar-based measurement model is proposed to update the proposal distribution which realizes the refine matching under severe occlusion in correction step in both global localization and pose tracking.

This paper is organized as follows. After discussing related work in the following section we introduce the theoretical basis. Then the localization system is described in detail, which presents the techniques of the landmark retrieval used to provide proposal distribution, and the newly proposed lidar-based measurement model used to estimate weight of particles; In Section V, various experiments are presented to illustrate the robustness of the system; Finally, we draw conclusions based on theoretical analysis and experimental results.

II. RELATED WORK

Compared to range sensors, cameras are low-cost sensors that provide a huge amount of information, so that vision-based navigation systems do not suffer from the interferences often observed when using light-based proximity sensors. Moreover, if robots are deployed in populated environments, it makes sense to base the perceptional skills used for localization on vision like humans do. Compared with vision, lidar is unparalleled in accuracy and efficiency in scale information, and it can serve as both observation information in localization and obstacle detection of navigation. As a result, the localization system based on vision and lidar fusion has been widely utilized.

A. LIDAR-BASED LOCALIZATION

Fox *et al.* proposed the Monte Carlo Localization [15], which applies sampling-based methods for approximating probability distributions. The number of samples is adapted on-line, thereby invoking large sample sets only when necessary, so it keeps the real-time and accuracy of pose tracking. Several methods for restoring global positioning are proposed for Monte Carlo Localization (MCL) failure in fatal state by Ueda *et al.* [16], where it is widely used to uniformly distribute particles throughout the space and then resample until convergence; The extended Kalman filter (EKF) was also used to localize the mobile robot with a laser range finder (LRF) sensor [17], dealing with the problem of estimating the output-noise covariance matrix that is involved in the localization of a mobile robot; Blanco *et al.* [18] enable the usage of the optimal proposal to estimate the true posterior density of a non-parametric dynamic system, which has greatly improved the efficiency of particle convergence; Geometrical FLIRT phrases (GFPs) was introduced as a novel retrieval method for efficient and precise place recognition [19], [20], they perform approximate 2D range data matching with low computational cost. However, opposed to visual information, in the environment of a single structure, geometric information retrieval relying on two-dimensional point clouds is still insufficient. Soonyong Park presented a coarse-to-fine global localization approach based on place learning with a 2-D range scan [21], coarse localization of the SVM-based place recognition selects candidate places where the robot may be located, and fine localization computes the relative poses to the candidate places with fast spectral scan matching and estimates the correct robot pose with a particle filter algorithm.

B. VISUAL-BASED LOCALIZATION

Wolf *et al.* [22] integrates an image retrieval system with Monte-Carlo localization, the image retrieval process is based on features that are invariant with respect to image translations and limited scale and then fine localization by lidar, and the system performs well in global localization and abduction recovery; Junqiu Wang *et al.* present a novel coarse-to-fine global localization approach inspired by object recognition and text retrieval techniques [23]. Harris–Laplace interest points characterized by scale-invariant transformation feature descriptors are used as natural landmarks, and epipolar geometry is used to refine localization. Nitsche presents a combination of a teach-and-replay visual navigation and Monte Carlo localization methods [24], robot can be started from at any point in the map and can deal with the 'kidnapped robot' problem. A coarse-to-fine strategy for global localization was suggested by [25], the coarse pose is estimated by means of object recognition and SVD-based point cloud fitting, and then is refined by stochastic scan matching; Naseer *et al.* [26] propose to use Markov localization-based temporal filtering over the similarity matrix to exploit the sequential information, but Markov-based visual positioning can only

provide relatively rough poses and cannot complete local refine matching; A method based on neural network-based visual retrieval and Monte Carlo localization is proposed [27]. It still uses the idea of coarse-to-fine, that is, coarse positioning through visual retrieval and fine matching with lidar.

The techniques described above either utilize visual-based landmark techniques or rely on simple features and use probabilistic state estimation to localize the robot, all achieved good localization results under ideal conditions. However, there is no solution for the localization failure caused by serious occlusion. What's more, any paper proposes a way to actively detect and record effective visual landmarks. The goal of this paper is to illustrate an approach by combining an image retrieval system with improved state estimation technique insensitive to severely occlusive environment. We describe how to actively record available landmarks and retrieve visual landmark via DBoW, and the improved lidar-based measurement model is applied to update the proposal distribution, finally the fine localization in the event of severe occlusion is completed. In practical experiments, we demonstrate that our approach is able to globally localize the robot and to reliably keep track of it when the lidar is seriously blocked.

III. THEORETICAL BASIS

Robot localization problems mainly include global localization and pose tracking. With subtle differences, the solutions to localization problems include the estimation of the prior of the robot poses up to the current instant of time given the whole history of available data. Let $x^t = \left\{ x^1, ..., x^t \right\}$ denote the sequence of robot poses up to time step t.

In the case of global localization, we are just interested in the current robot pose instead of the whole path following the rule in (1), where υ^t is the observation of camera which will provide prior information, and z^t represent the observation of lidar which will update the prior.

$$p\left(x^t | z^t, \upsilon^t\right) \propto \overbrace{p\left(z_t | x^t, \upsilon^t\right)}^{Observation\ likelihood} \overbrace{p\left(x^t | \upsilon^t\right)}^{Prior} \quad (1)$$

In the case of pose tracking, the posterior of the robot pose can be computed sequentially by applying the Bayes rule in (2), where the z^t and the u^t represent the sequences of robot observations and actions.

$$p\left(x^t | z^t, u^t\right) \propto \overbrace{p\left(z_t | x^t, u^t\right)}^{Observation\ likelihood} \overbrace{p\left(x^t | z^{t-1}, u^t\right)}^{Prior} \quad (2)$$

The belief of robot's posture is represented by a set of random samples, and the Monte Carlo Localization is implemented to update the belief. As shown in Algorithm 1, we denote χ_{t-1}, χ_t as particle sets at different time, S is the number of particles. In prediction phase, the robot performs sampling based on the prior information provided by the *sample_motion_model* to generate a proposal distribution, where μ_t represent the transition caused by motion control; In correction phase, the weight of particles is updated based

---

Algorithm 1 MCL

Input: $(\chi_{t-1}, \mu_t, z_t, m)$
1: **for** $s = 1\ to\ S$ **do**
2:     $x_t^{[s]} = sample\_motion\_model(\mu_t, x_{t-1}^{[s]})$
3:     $\omega_t^{[s]} = measurement\_model(z_t, x_t^{[s]}, m)$
4:     $\bar{\chi}_t = \bar{\chi}_t + \left\langle x_t^{[s]}, \omega_t^{[s]} \right\rangle$
5: **end for**
6: **for** $s = 1\ to\ S$ **do**
7:     draw s with probability $\propto \omega_t^{[s]}$
8:     add $x_t^{[s]}$ to χ_t
9: **end for**
10: return χ_t

---

on *measurement_model*, where z_t represents the observation model of the ranging sensor, m represents the grid map information. Resampling performed from line 6 to line 9 is to avoid particle degradation.

In this paper, the prediction step and correction step are improved respectively. When the robot is in global localization, the prediction step generates a proposal distribution based on the prior information provided by visual retrieval following the (1). The correction step is performed using the lidar-based measurement model proposed in this paper. Instead of the traditional *measurement_model* that the laser beams are used at equal intervals for particle weight evaluation, the improved *measurement_model* use the same number of laser beams with maximum weight to update the particles, because the larger the weight of beam is, the higher confidence the observation information is. This prevents the disturbed laser beams from updating the particles as observation, thereby ensuring the localization robustness of the robot under severe occlusion. During the pose tracking, odometer motion model is utilized to generate the proposal distribution following the (2), and the improved *measurement_model* is used for correction step.

IV. SYSTEM OVERVIEW AND METHODOLOGY
A. SYSTEM OVERVIEW
On the basis of theory in section III, a new robot localization system is designed. Firstly, a map with a visual landmark library was built. Then the visual information searched from the landmark library was used for proposal distribution, and finally, the improved lidar-based measurement model performs local refine matching. The entire system is divided into two parts, the off-line stage and the online stage, as shown in Figure 1.

The off-line stage mainly constructs map with visual landmarks and generates off-line visual DBoW. Simultaneous SLAM localization and mapping (SLAM) based on RBPF is adopted to map and pose estimation, and the Oriented FAST and Rotated BRIEF (ORB) features are extracted from obtained pictures for matching, and the matching rate proposed in this paper is regarded as the criterion to select the valid landmarks, at the same time, pose information estimated
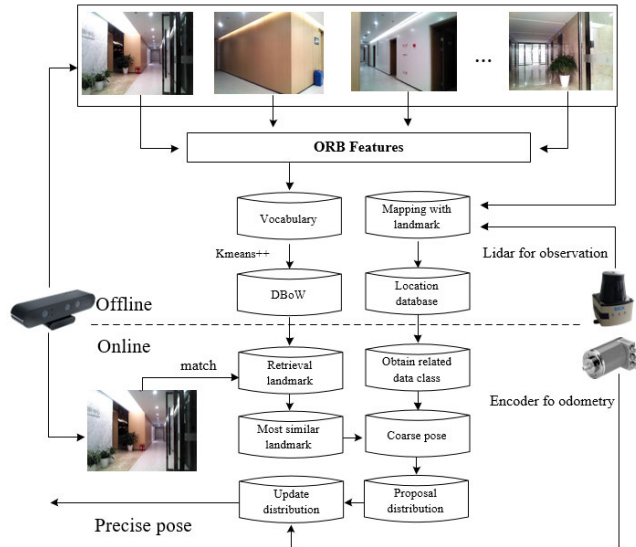
FIGURE 1. System overview.

FIGURE 2. Active landmark system.

by SLAM associated with the landmarks is also recorded (Where the robot sees the landmarks). Finally, the ORB features are extracted from the recorded landmarks, and kmeans++ training is used to generate an off-line DBoW to prepare for the landmarks retrieval in the online stage.

The online stage is divided into two phases of prediction and correction. In prediction phase of global localization, the DBoW based search algorithm is adopted to obtain the closest to the current scene, and the robot's rough pose information corresponding to the most similar landmark is to generate proposal distribution of particle set, while the prediction phase of pose tracking use odometer from encoder as the prior information. In correction stage of both global localization and pose tracking, the newly proposed lidar-based measurement model is applied to update particles to complete refine localization.

B. OFF-LINE STAGE
The mapping process in this paper uses the SLAM algorithm based on RBPF, which uses the odometer model as the prior information to generate the proposal distribution, scan matching process are carried out to maximum the likelihood function by a Gaussian to update robot proposal distribution. While mapping, the ORB feature is used to detect and match key points of the image, and key frames are selected and recorded according to certain rules. The robot continuously publishes pose information during the mapping process, and the landmark detection system records valid landmarks with the pose information of the robot. The process is shown in Figure 2.

Landmarks should have the following characteristics:(a) the pictures should have enough feature information, and the feature extraction process should be of highly real-time performance, which is convenient for fast matching; (b)we should make a trade-off between amount of spatial
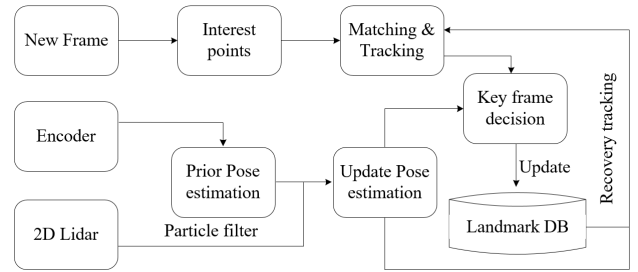
information and number of landmarks, and use as few landmarks as possible to represent the densest spatial location information; (c) we should avoid interruption of the recording system due to tracking lost. In summary, the landmark recording process in this paper is carried out according to the following rules:

(1) Landmark detection performs every 5 frames;
(2) ORB features are extracted for each frame, and the frames with more than 200 key points are selected as the key frame according to experience value;
(3) The matching rate between the current frame with the previous key frame is in a threshold range, and the empirical value is $0.1 \sim 0.55$;
(4) If the matching rate of key frames for 4 consecutive frames is not within the threshold, it is considered to be tracking lost, and then the process continues after the current landmark is recorded.

Rule (1) and rule (3) meet the requirement of (b), avoiding repeated recording of landmarks when stationary or excessively overlapping landmarks recorded during movement, and as much as possible to ensure a trade-off between feature information and the number of landmarks. ORB is rotation invariant and resistant to noise, ORB is at two orders of magnitude faster than Scale-invariant feature transform (SIFT) [28]. Therefore, the ORB is used to extract the visual features, and the rule (2) is used to meet the requirement (a). At the same time, the concept of matching rate is proposed to assist in the selection of effective landmarks, avoiding excessive overlapping of landmarks, causing database explosion, and the matching rate is calculated as formula (3) shown, where ε represents the matching rate, α is the number of matching key points between the prior frame and the current frame, λ_k and λ_{k-1} respectively indicate the number of key points of the current key frame and the previous key frame.

$$\varepsilon = 2 \times \alpha/(\lambda_k + \lambda_{k-1}) \qquad (3)$$

The matching result within the threshold of matching rate is shown in Figure 3; Tracking loss is easy to occur when the robot performs landmark detection and recording during the movement. As shown in Figure 4, the direct cause of tracking loss is the sharp increase in the number of key points of mismatches, which causes the matching rate beyond the threshold consecutively. In order to avoid the situation
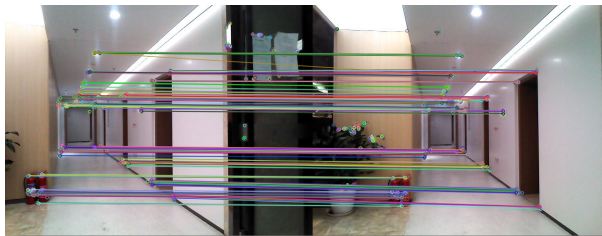
FIGURE 3. Tracking successful ($\varepsilon = 0.428$).

FIGURE 4. Tracking lost ($\varepsilon = 0.825$).

that the recording system break off due to tracking lost, in rule (3), the robot determines whether it is in the tracking lost state according to the matching rate. Once in the lost state, it records the current frame and continues tracking and recording to ensure the continuity of recording system, which meets the requirement of (c).

According to the above method, a probability grid map as shown in Figure 5 is generated, and the position distribution of the recorded landmarks is shown in Figure 6. It can be seen that the landmarks cover almost the entire grid map, which represents relatively dense space information with sparse visual features. Then the recorded landmarks are trained to generate a DBoW. The DBoW is a technique that allows to convert with a visual vocabulary a set of local features coming from an image into a sparse numerical vector, allowing to manage big sets of images [13]. In order to retrieval similar landmark we use an image database composed of a hierarchical bag of words, and direct and inverse indexes. The visual vocabulary is created by discretizing the descriptor space into several visual words. We discretize a binary descriptor space by ORB, creating a more compact vocabulary. The vocabulary tree is built by kmeans++ algorithm as proposed in [29]. Every word or leaf node maintained an inverse index, it stores a list of images where the word present. We weight each word with its inverse document frequency (*idf*), where is the number of training images, and the number of occurrences of word in these images.

$$idf(i) = \log \frac{N}{n_i} \qquad (4)$$

When querying the similar landmarks, image I is converted into a bag-of-words vector $\upsilon = \{\omega_1, \omega_2, ..., \omega_m\}$, the binary descriptors of its m features traverse the tree from the root to the leaves, by selecting at each level the intermediate nodes that minimize the Hamming distance. This allows us to
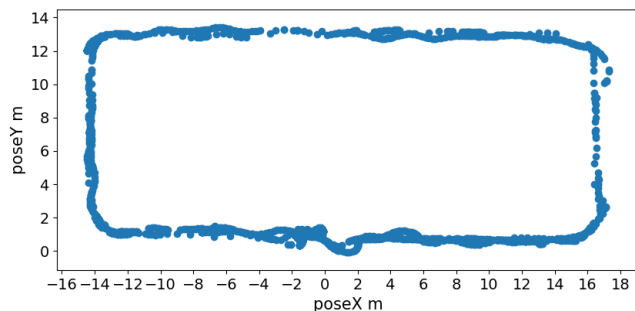

FIGURE 5. Grid map.

FIGURE 6. Landmark location in grid map.

calculate the term frequency (*tf*) of each word in this image.

$$f(i, I) = \frac{n_{iI}}{n_I} \qquad (5)$$

n_{iI} stands for the number of occurrences of word η_i in image I, and n_I for the number of words in I. The i-th entry of υ is finally given the value $\eta_t^i = tf(i, I_t) \times idf(i)$, obtaining the $tf - idf$ weight. The words corresponding to the key points of each image constitute the bag-of-words vector v of this image:

$$v = \{(\omega_1, \eta_1), (\omega_2, \eta_2), ..., (\omega_m, \eta_m)\} \qquad (6)$$

To measure the similarity between two landmarks v_A and v_B, we can calculate a L1-score $s(v_A - v_B)$, when the database is queried, all the matches are ranked by their scores and the best picture will be selected.

$$s(v_A - v_B) = 2 \sum_{i=1}^{m} \left|v_{A_i}\right| + \left|v_{B_i}\right| - \left|v_{A_i} - v_{B_i}\right| \qquad (7)$$

C. ONLINE STAGE

In general, Monte Carlo-based global localization uses the random selection of an initial position on map to uniformly distribute a large number of particles on the map to generate the proposal distribution [16]. The weight of particles continuously updates according to the lidar-based measurement model until it converges to the exact position in correction phase. Once the initial position is far away from the actual position, it is of great possibility to cause localization failure. In the correction phase, the probability of effective observation is equal to that of invalid or even harmful observation, because the lidar measurement model samples the laser
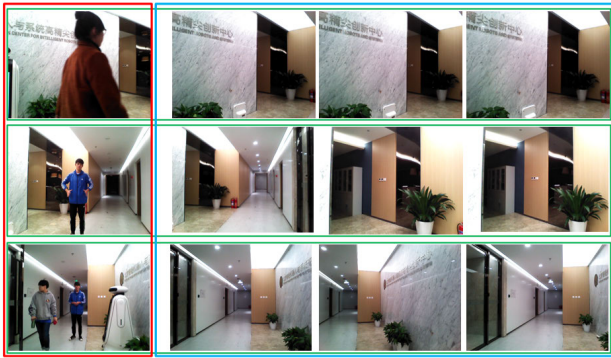
FIGURE 7. Retrieval result by DBoW: in the red box are pictures observed by the robot at different positions, in the blue box are landmarks, and in the green box are each observation picture and the corresponding three most similar pictures (for interpretation of the references to color in this figure legend, the reader is refered to the web version of this article).

beams uniformly. Once the lidar is obscured severely, it is not possible to make full use of effective measurements to update the particles and resulting in localization failure.

In this paper, both the prediction phase and correction phase are improved respectively as discussed in section III. In the prediction phase of global localization, the proposal distribution is determined according to the error model of visual retrieval, while the odometer provides a reliable priori in the prediction phase of pose tracking. The improved lidar-based measurement model that the laser beams with high confidence are selected as the observation is adopted in the correction step of both global localization and pose tracking for fine localization.

Prediction phase: DBoW generated in off-line stage is applied to retrieve the most similar landmark, the robot determines the particle distribution according to the position information and the error model of the most similar landmark. Search results shown in Fig. 7 indicate that the retrieved landmark via DBoW under occlusion can also find corresponding landmarks. The highest-scoring landmark is to provide prior information to generate proposal distribution of particles.

In order to further determine the particle distribution of the initial position, the error model of the retrieved landmark is analyzed. Landmarks are clustered according to matching rate firstly, as shown in Algorithm 2. We denote $C = C_0, C_1, \ldots, C_L$ as the set of landmark classes, $C_l = P_n, P_{n+1}, \ldots (l \in (0, L))$ represents the landmark class, L is the number of classes, and it can only be determined after clustering has been completed, P_n is denoted as landmark, N is the maximum number of landmarks, *threshold* $\in$ [0.1, 0.55] based on experience. When the matching rate between the current picture P_n and the next picture P_{n+1} is within the threshold range, P_{n+1} will be added to the class of the current picture, otherwise, a new class will be created and P_{n+1} will be added to the new class until all the pictures are traversed to complete the clustering.
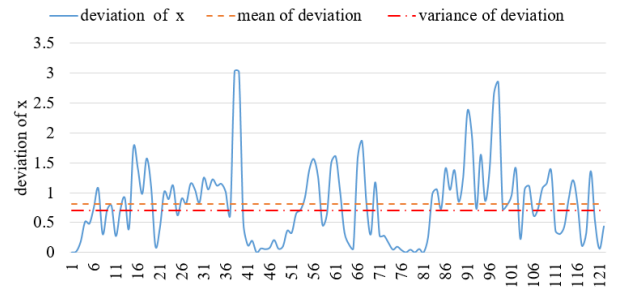
After the clustering is completed, the mean values of the positions and postures associated with various landmarks

---

Algorithm 2 Landmark Clustering

1: $C = \emptyset, C_l = \emptyset, l = 0$
2: **for** all landmarks $P_n = P_0$ **to** P_N **do**
3:      **if** match(P_n, P_{n+1}) $\in$ *threshold* **then**
4:          $C_l = C_l \cup P_n, P_{n+1}$
5:      **else**
6:          $C = C \cup C_l$
7:          $l + +$ (create new C_l)
8:          $n + +$
9:      **end if**
10: **end for**
11: return C

FIGURE 8. Error distribution, mean and variance of adjacent classes in X.

belong to the classes are taken as the poses of the landmark classes. Owing to the error range between the landmarks is smaller than the error range between the landmarks, it is feasible to analyze the rough distribution of the robot's proposal distribution of global position using the error model between the landmarks. According to Figure 8–10, the variance σ_{xx}, σ_{yy}, σ_{zz} of adjacent landmark classes in different dimensions of x, y, and z (yaw) can be obtained. When the robot retrieves a picture from the landmark database that is most similar to the current scene, the particles representing the robot's pose should obey the Gaussian distribution with the position of the landmark class where the picture is located as the mean and Σ as the covariance matrix, as shown in formula (8), where $\mu = p(x, y, z)$ is location of class that the most similar landmark belong to, $\Sigma = diag(\sigma_{xx}, \sigma_{yy}, \sigma_{zz})$.

$$\mathcal{N}(p|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}}$$
$$\times \exp\left\{-\frac{1}{2}(p - \mu)^T \Sigma^{-1}(p - \mu)\right\} \quad (8)$$

In order to maximize the confidence of the particle distribution range, according to the 3Σ rule in formula (9), covariance matrix of the particle distribution is set as $\Sigma = diag\left(\sigma'_{xx}, \sigma'_{yy}, \sigma'_{zz}\right)$, where $\sigma'_{xx} = (\sqrt{\sigma_{xx}} \times 3)^2$, $\sigma'_{yy} = (\sqrt{\sigma_{yy}} \times 3)^2$, $\sigma'_{zz} = (\sqrt{\sigma_{zz}} \times 3)^2$.

$$Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997 \quad (9)$$

In this paper, $\Sigma = diag(7.02, 3.42, 1.08)$.

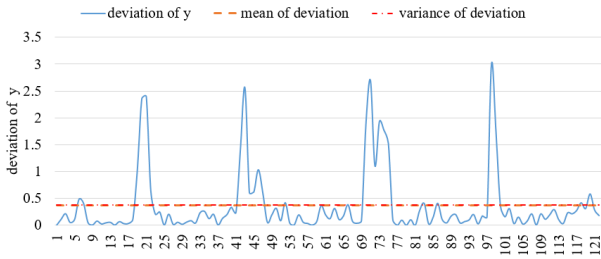**Correction phase:** The measurement model proposed in this paper selects the laser beams with high confidence

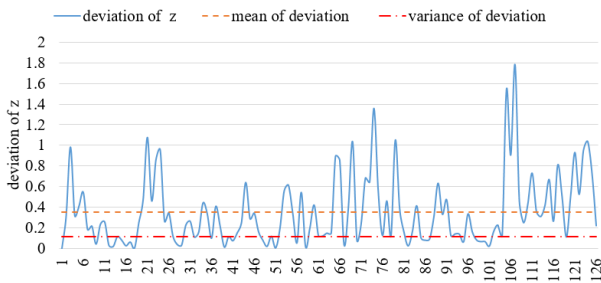**FIGURE 9.** Error distribution, mean and variance of adjacent classes in Y.

FIGURE 10. Error distribution, mean and variance of adjacent classes in Z (yaw).

Algorithm 3 Measurement Model

Input: (z_t, x_t, m)

1: **for** $s = 1$ **to** S **do**
2:    **for** $(k = 0;\ k < laser\_count;\ k+ = step)$ **do**
3:       $p = z_{hit} \cdot p_{hit}(z_t^k|x_t, m) + z_{short} \cdot p_{short}(z_t^k|x_t, m) + z_{max} \cdot p_{max}(z_t^k|x_t, m) + z_{rand} \cdot p_{rand}(z_t^k|x_t, m)$
4:       $weight$.push_back($p*$)
5:    **end for**
6:    sort the elements of $weight$ from big to small
7:    **for** $(i = 0;\ i < using\_laser\_count;\ i + +)$ **do**
8:       $p+ = weight(i)$
9:    **end for**
10:   $\omega_t^{[s]} = \omega_t^{[s]} \times p$
11:   $total\_weight+ = \omega_t^{[s]}$
12: **end for**
13: **return** $total\_weight$

---

as the observation to evaluate the weight of particles. The *measurement_model* is shown in Algorithm 3, where *laser_count* is the total number of laser beams of lidar, *step* represents sampling step of laser beams, *using_laser_count* is denoted as the number of laser beams for observation, and the following conditions need to be met.

$$laser\_count/step \geq using\_laser\_count \qquad (10)$$

We denote $z_t = \{z_t^1, z_t^2, \cdots, z_t^k\}$ as the measurement of all beams, x_t is the pose state, p is the observation likelihood of each beam, $\omega_t^{[s]}$ is weight for each particle, the *total_weight* is the total weight of the particles, which is used to normalize the particle weights in resampling process.

The algorithm from line 2 to line 6 is to calculate the likelihood of each laser beam, and the

using_laser_count beams with the highest confidence are selected. *Observation_likelihood* in line 3 is represented by *beam_model* [6], which incorporates four types of measurement errors, all of which are essential to making this model work: small measurement noise $p_{hit}(z_t^k|x_t, m)$, errors due to unexpected objects $p_{short}(z_t^k|x_t, m)$, errors due to failures to detect objects $p_{max}(z_t^k|x_t, m)$, this four different distributions are weighted by 4 parameters to calculate the lidar's observational likelihood on the map, and $z_{hit} + z_{short} + z_{max} + z_{rand} = 1$ as is shown in formula (11).

$$p(z_t^k|x_t, m) = z_{hit}p_{hit} + z_{short}p_{short} + z_{max}p_{max}$$
$$+ z_{rand}p_{rand} \qquad (11)$$

Instead of the traditional algorithm that the laser beams are selected by uniformly-spaced sampling, the laser beams with highest confidence are selected as observation to update weight of particles and the confidence levels of the laser beams are sorted according to the observation likelihood (line 6). In this way, when the lidar is blocked by a large area, both static and dynamic obstacle information, as long as the obstacle information is unreliable, they can avoid being introduced to update particles, so as to use observation information with high confidence to update the proposal distribution, and the robustness localization is enhanced in the case of serious occlusion. With this method, although the number of laser beams per sampling is higher than that of the traditional method, the improved measurement model makes the number of iterations and particles in the process of particle filter greatly reduced, which improves the efficiency and accuracy of localization.

V. RESULTS AND ANALYSIS

Experiments are conducted for global localization and pose tracking under severe occlusion. Since the global localization process has been improved from the prediction phase and the correction phase respectively, four groups of experiments based on the idea of control variables are carried out as follows to verify the superiority of the localization system in this paper, and all experiments are carried out under the relatively harsh conditions, which is that the maximum number of particles is 300, the number of 1/12 laser beams is used as observation, and 60% excitation is under the occlusion condition. FL in the following figure refers to the fake lidar data returned by the occluded lidar, CL refers to the correct lidar value, AP refers to the actual position of the robot (the Green Point), EP is the robot estimated position according to the localization algorithm (yellow point), and the blue point represents the particles. When AP and EP coincide, the robot localization is successful and the particles converge to the distribution range near the actual position, otherwise, the localization failed. In pose tracking process, the odometer is used as the prior information in the prediction phase, and the measurement model proposed in this paper is used in the correction phase. The trajectory of the odometer and the trajectory calibrated by two localization algorithms are recorded for comparison.
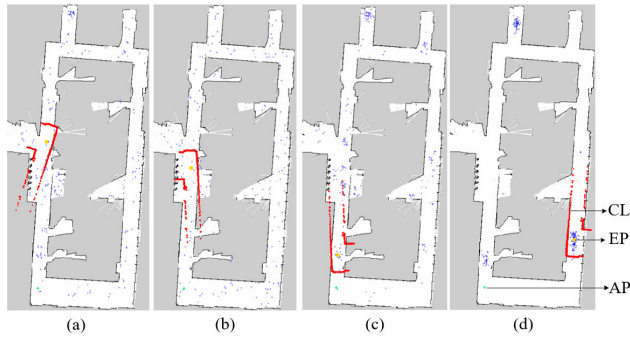
FIGURE 11. Global localization process without occlusion of traditional prediction model and measurement model, the particle filter iterates 194 times and costs 7.1s.
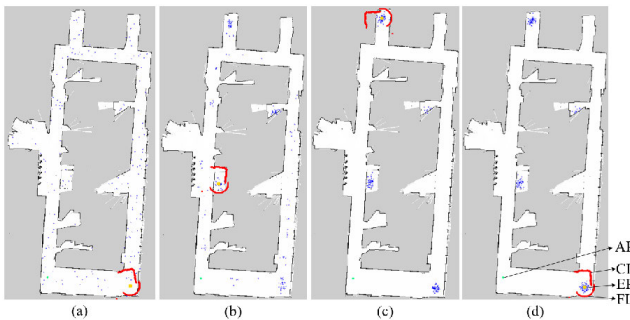
FIGURE 12. Global localization process under occlusion of traditional prediction model and measurement model, the particle filter iterates 51 times and costs 1.7s.
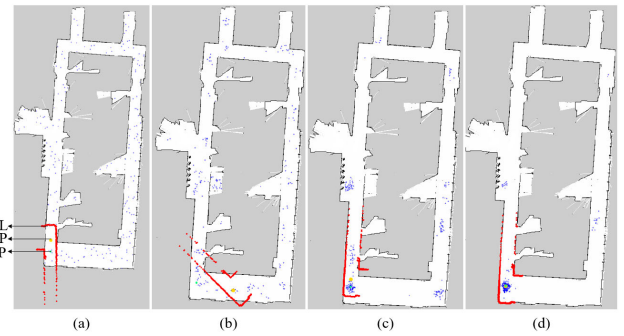

FIGURE 13. Global localization without occlusion of improved measurement model and traditional prediction, the particle filter iterates 44 times and costs 6.4s.

FIGURE 14. Global localization without occlusion of improved measurement model and traditional prediction, the particle filter iterates 41 times and costs 6.2s.

A. GLOBAL LOCALIZATION

The prediction model and measurement model of global localization are both improved. The initial position is randomly drawn in grid map in traditional prediction model, but the improved prediction model provide the distribution of initial position by visual retrieval. The traditional measurement model does not distinguish between reliable and unreliable observation, while the improved measurement model will select reliable lidar beams as observation.

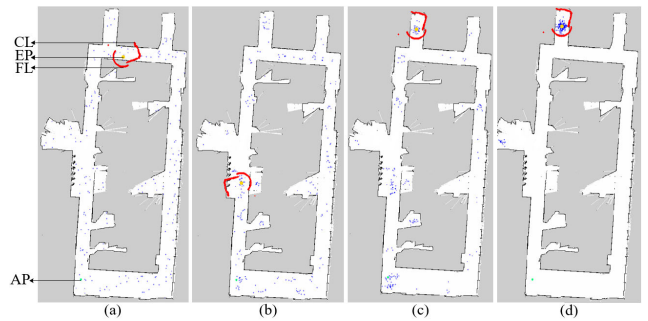#### 1) TRADITIONAL PREDICTION MODEL + TRADITIONAL MEASUREMENT MODEL

In the traditional AMCL global localization process, the initial position is set randomly and the particles are distributed evenly for iterative calculation. Particles converges to the wrong position under both normal and severe occlusion conditions, resulting in positioning failure, as is shown in Figure 11 and Figure 12.

2) TRADITIONAL PREDICTION MODEL + IMPROVED MEASUREMENT MODEL

By using the improved measurement model, the particles are evenly distributed in the global environment for iterative calculation. It can be seen that when the lidar is not occluded, the robot converges to the correct position as is shown in Figure 13, so it can be proved that the improved measurement model is better than the original measurement model; when the lidar is seriously occluded, the robot converges to the wrong position as is shown in Figure 14, which indicates that

there are some limitations in the global localization by lidar alone.

3) IMPROVED PREDICTION MODEL+ TRADITIONAL MEASUREMENT MODEL

When using the improved prediction method and traditional measurement model for global localization, due to the situation that the predicted position is far from the real position (0.2 m in x direction, 6 m in y direction, 0.3 rad in z direction), and the results of erroneous convergence occur in both occluded and non-occluded cases as shown in Figure 15 and Figure 16, so it proves that the traditional AMCL is still not strong enough when weak prior information is provided.
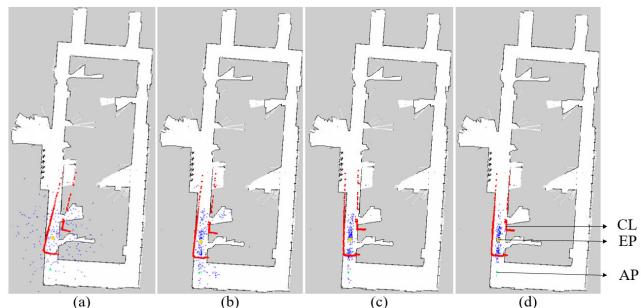
4) IMPROVED PREDICTION MODEL+ IMPROVED MEASUREMENT MODEL

After adopting improved prediction model and improved measurement model, the robot can converge to an accurate position with or without occlusion as shown in Figure 17 and Figure 18, which proves the superiority of this system.
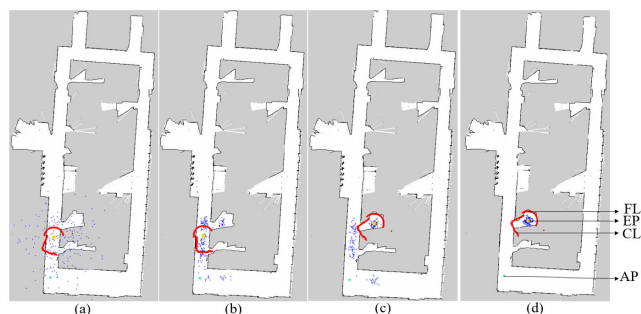
B. POSE TRACKING

After completing the global localization, it is also necessary to meet the real-time and accuracy of pose tracking during

FIGURE 15. Global localization without occlusion of traditional measurement model and improved prediction model, the particle filter iterates 20 times and costs 1.1s.
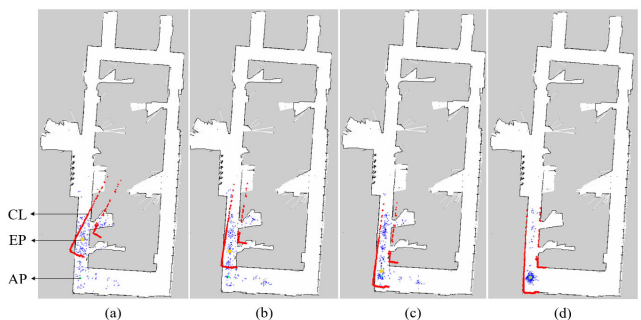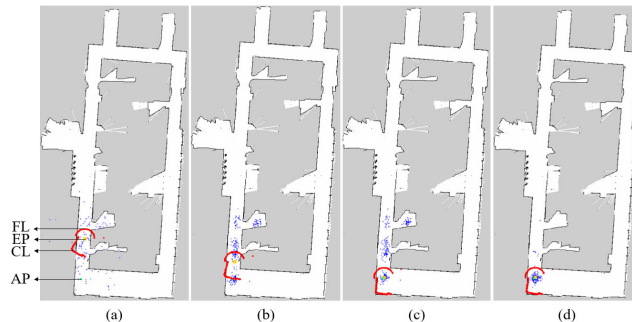

FIGURE 16. Global localization with occlusion of traditional measurement model and improved prediction model, the particle filter iterates 52 times and costs 2.8s.
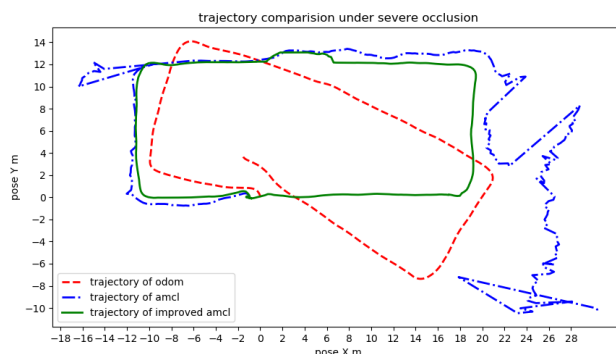

FIGURE 17. Global localization without occlusion of improved measurement model and improved prediction model, the particle filter iterates 21 times and costs 3.0s.

FIGURE 18. Global localization with occlusion of improved measurement model and improved prediction model, the particle filter iterates 27 times and costs 3.4.

FIGURE 19. Trajectories of odometer, AMCL and improved AMCL.

the robot's continuous movement. In order to verify the performance of the improved algorithm, during the experiment, the robot with a lidar blocking 60% moves at a speed of 1.2m/s, odometer trajectory and the tracking trajectory output by the two localization algorithms are recorded. AMCL still uses the original measurement model, and the algorithm of this paper uses the improved measurement model.

As shown in Figure 19, the overall trajectory of the odometer is deformed due to the accumulated error, and the AMCL outputs a wrong trajectory because the lidar is severely blocked. However, under the same motion scenario, the improved AMCL with improved measurement model corrects the odometer data in real time to ensure the accuracy of robot positioning.

VI. CONCLUSION

In this paper, a robust localization system is designed to solve the problem of robot location failure when lidar is blocked seriously. Based on Monte Carlo positioning, the system improves the prediction and correction of the robot. In the prediction stage, the offline DBoW generated by training is used to retrieve similar landmarks, and the proposal distribution is generated according to the retrieved visual landmarks; in the update stage, the measurement model of lidar is improved, and the laser beams with higher confidence are selected as the observation to update particles, which avoids the observation noise caused by the occlusion of the introduced pseudo data, and improves the robustness of positioning. Experimental results show that compared with the state of art non parametric filtering algorithm AMCL, the system has obvious advantages in global localization and pose tracking. In future work, a landmark system incorporating semantic information will be used to improve the accuracy of visual retrieval in the case of drastic changes in lighting, color, and occlusion, so as to provide a reliable priority for the global localization of robot.

REFERENCES

[1] X. Qu, B. Soheilian, and N. Paparoditis, "Landmark based localization in urban environment," *ISPRS J. Photogramm. Remote Sens.*, vol. 140, pp. 90–103, Jun. 2018.

[2] A. Amanatiadis, "A multisensor indoor localization system for biped robots operating in industrial environments," *IEEE Trans. Ind. Electron.*, vol. 63, no. 12, pp. 7597–7606, Dec. 2016.

[3] Y. Xu, Y. S. Shmaliy, Y. Li, X. Chen, and H. Guo, "Indoor INS/LiDAR-based robot localization with improved robustness using cascaded FIR filter," *IEEE Access*, vol. 7, pp. 34189–34197, 2019.

[4] F. Boniardi, T. Caselitz, R. Kummerle, and W. Burgard, "Robust LiDAR-based localization in architectural floor plans," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 3318–3324.

[5] J. Rowekamper, C. Sprunk, G. D. Tipaldi, C. Stachniss, P. Pfaff, and W. Burgard, "On the position accuracy of mobile robot localization based on particle filters combined with scan matching," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 3158–3164.

[6] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA, USA: MIT Press.

[7] L. Jetto, S. Longhi, and G. Venturini, "Development and experimental validation of an adaptive extended Kalman filter for the localization of mobile robots," *IEEE Trans. Robot. Autom.*, vol. 15, no. 2, pp. 219–229, Apr. 1999.

[8] K. Jung, J. Kim, J. Kim, E. Jung, and S. Kim, "Positioning accuracy improvement of laser navigation using UKF and FIS," *Robot. Auto. Syst.*, vol. 62, no. 9, pp. 1241–1247, Sep. 2014.

[9] L. D'Alfonso, W. Lucia, P. Muraca, and P. Pugliese, "Mobile robot localization via EKF and UKF: A comparison based on real data," *Robot. Auto. Syst.*, vol. 74, pp. 122–127, Dec. 2015.

[10] Z. Chen, "Bayesian filtering: From Kalman filters to particle filters, and beyond," *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

[11] D. Fox, "Adapting the sample size in particle filters through KLD-sampling," *Int. J. Robot. Res.*, vol. 22, no. 12, pp. 985–1003, Dec. 2003.

[12] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with rao-blackwellized particle filters," *IEEE Trans. Robot.*, vol. 23, no. 1, pp. 34–46, Feb. 2007.

[13] D. Galvez-Lopez and J. D. Tardos, "Real-time loop detection with bags of binary words," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 51–58.

[14] D. Galvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Trans. Robot.*, vol. 28, no. 5, pp. 1188–1197, Oct. 2012.

[15] D. Fox, W. Burgard, F. Dellaert, and S. Thrun, "Monte Carlo localization: Efficient position estimation for mobile robots," in *Proc. AAAI/IAAI*, 1999, nos. 343–349, p. 2.

[16] R. Ueda, T. Arai, K. Sakamoto, T. Kikuchi, and S. Kamiya, "Expansion resetting for recovery from fatal error in monte carlo localization-comparison with sensor resetting methods," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, vol. 3, Sep./Oct. 2004, pp. 2481–2486.

[17] L. Teslić, I. Škrjanc, and G. Klančar, "Using a LRF sensor in the Kalman-filtering-based localization of a mobile robot," *ISA Trans.*, vol. 49, no. 1, pp. 145–153, Jan. 2010.

[18] J.-L. Blanco, J. González, and J.-A. Fernández-Madrigal, "Optimal filtering for non-parametric observation models: Applications to localization and SLAM," *Int. J. Robot. Res.*, vol. 29, no. 14, pp. 1726–1742, Dec. 2010.

[19] G. D. Tipaldi, L. Spinello, and W. Burgard, "Geometrical FLIRT phrases for large scale place recognition in 2D range data," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2013, pp. 2693–2698.

[20] V. Ratziu, P. Giral, S. Jacqueminet, F. Charlotte, A. Hartemann-Heurtier, L. Serfaty, P. Podevin, J. Lacorte, C. Bernhardt, E. Bruckert, A. Grimaldi, and T. Poynard, "Rosiglitazone for nonalcoholic steatohepatitis: One-year results of the randomized placebo-controlled fatty liver improvement with rosiglitazone therapy (FLIRT) trial," *Gastroenterology*, vol. 135, no. 1, pp. 100–110, Jul. 2008.

[21] S. Park and K. S. Roh, "Coarse-to-Fine localization for a mobile robot based on place learning with a 2-D range scan," *IEEE Trans. Robot.*, vol. 32, no. 3, pp. 528–544, Jun. 2016.

[22] J. Wolf, W. Burgard, and H. Burkhardt, "Robust vision-based localization by combining an image-retrieval system with Monte Carlo localization," *IEEE Trans. Robot.*, vol. 21, no. 2, pp. 208–216, Apr. 2005.

[23] J. Wang, H. Zha, and R. Cipolla, "Coarse-to-fine vision-based localization by indexing scale-invariant features," *IEEE Trans. Syst., Man Cybern., B (Cybern.)*, vol. 36, no. 2, pp. 413–422, Apr. 2006.

[24] M. Nitsche, T. Pire, T. Krajník, M. Kulich, and M. Mejail, "Monte Carlo localization for teach-and-repeat feature-based navigation," in *Advances in Autonomous Robotics Systems*, M. Mistry, A. Leonardis, M. Witkowski, and C. Melhuish, Eds. Cham, Switzerland: Springer, 2014, pp. 13–24.

[25] S. Park, S. Kim, M. Park, and S.-K. Park, "Vision-based global localization for mobile robots with hybrid maps of objects and spatial layouts," *Inf. Sci.*, vol. 179, no. 24, pp. 4174–4198, Dec. 2009.

[26] T. Naseer, B. Suger, M. Ruhnke, and W. Burgard, "Vision-based Markov localization for long-term autonomy," *Robot. Auto. Syst.*, vol. 89, pp. 147–157, Mar. 2017.

[27] S. Xu, W. Chou, and H. Dong, "A robust indoor localization system integrating visual localization aided by CNN-based image retrieval with Monte Carlo localization," *Sensors*, vol. 19, no. 2, p. 249, 2019.

[28] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. ICCV*, Nov. 2011, vol. 11, no. 1, p. 2.

[29] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1465–1479, Sep. 2006.

YONGLIANG SHI received the B.S. and M.S. degrees from the North University of China, in 2013. He is currently pursuing the Ph.D. degree in mechanics engineering with the Beijing Institute of Technology. His research interests include SLAM and semantic localization for robotics.

WEIMIN ZHANG received the B.S., M.S., and Ph.D. degrees in mechatronics engineering from the Beijing Institute of Technology (BIT), Beijing, China, in 1999, 2002, and 2005, respectively. He was a Visiting Scholar with the Department of Modern Mechanical Engineering, Waseda University, Japan, in 2008. He is currently an Associate Professor with the School of Mechatronical Engineering, BIT. His research interests include mechanical design and bionic vision of humanoid robot.

FANGXING LI received the B.S. degree from the Chiba Institute of Technology, in 2001, the M.S. degree from Chuo University, in 2004, and the Ph.D. degree from the Beijing Institute of Technology (BIT). He is currently a Lecturer with the School of Mechatronical Engineering, BIT. His research interests include mechanical design and bionic vision of robot navigation.

QIANG HUANG received the B.S. and M.S. degrees in electrical engineering from the Harbin Institute of Technology, China, in 1986 and 1989, respectively, and the Ph.D. degree in mechanical engineering from Waseda University, Japan, in 1996. He was a Research Fellow with the National Institute of Advanced Industrial Science and Technology, Japan, from 1996 to 1999. He was a Research Fellow with The University of Tokyo, Japan, from 1999 to 2000. He is currently a Professor at BIT. He is also the Director of the Intelligent Robotics Institute and the Key Laboratory of Biomimetic Robots and Systems, Ministry of Education of China. His research interests include biped locomotion and bio-robotic systems. He received the Ministry of Education Award (First Class Prize) for Technology Invention. He served as the chairs for many IEEE conferences, such as the Organizing Committee Chair of the 2006 IEEE/RSJ IROS and the General Chair of the 2017 IEEE ROBIO and the 2018 IEEE-RAS ICHR.

● ● ●