

Received November 25, 2019, accepted December 16, 2019, date of publication December 25, 2019, date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962236

# Optimized Data Association Based on Gaussian Mixture Model

XIAOGANG RUAN<sup>1,3</sup>, DINGQI REN<sup>1,3</sup>, XIAOQING ZHU<sup>1,2,3</sup>, AND SHAODA LIU<sup>1,3</sup>

<sup>1</sup>Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

<sup>2</sup>College of Electronic and Information, Nanchang Institute of Technology, Nanchang 330099, China

<sup>3</sup>Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China

Corresponding author: Xiaoqing Zhu (alex.zhuxq@bjut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773027, in part by the Natural Science Foundation of Beijing under Grant 4202005, and in part by the Project of S&T Plan of Beijing Municipal Commission of Education under Grant KM201810005028.

**ABSTRACT** Data association is the foundation of state estimation in mobile robot simultaneous localization and mapping. Aiming at the problems of false association, high computational complexity in joint compatible branch and bound algorithm, we propose an optimized joint compatible branch and bound data association algorithm based on Gaussian mixture clustering. Firstly, the local association strategy is adopted to limit data association in local region, so as to reduce the number of features involved in data association at the current moment. Secondly, the Gaussian mixture clustering algorithm is used in local areas to group the observed values at the current moment, so as to get several groups that have little correlation with each other. Finally, joint compatible branch and bound data association algorithm is used in each group for data association, and the optimal solution is obtained according to mutual exclusion criteria and optimal criteria. The experiment results verify that the algorithm improved the accuracy of data association, reduced the computational complexity and improved the efficiency of data association.

**INDEX TERMS** Artificial intelligence, autonomous agents, clustering algorithms, intelligent robots, simultaneous localization and mapping.

## I. INTRODUCTION

Simultaneous localization and mapping (SLAM) technology is the key to realize the robot navigation. It means of a mobile robot in an unknown environment constantly sense the environment using its own sensor starting from the starting point. According to the sensor data to estimate its own position for achieving self-positioning [1]–[4]. Meanwhile, the environment map is built incrementally on the basis of localization to achieve its positioning and navigation. There are several solutions proposed to solve SLAM problems, including laser SLAM like GMapping [5], and LOAM [5], [6], visual SLAM like LSD-SLAM [7], ORB-SLAM2 [8], and RTAB-MAP [6], [9], semantic SLAM like DS-SLAM [7], [10], and DynaSLAM [11]. Data association is a key problem of mobile robot SLAM, which is also known as consistency problem.

Specifically, the purpose of data association is to determine whether there is a correspondence between the measured

The associate editor coordinating the review of this manuscript and approving it for publication was Ran Cheng.

values obtained by the mobile robot and the map features in the existing environment map under different time and environment region. And it determines whether the measured values come from the same entity in the environment based on the above correspondence. A few data association failures will cause the algorithm to diverge. Therefore, it is vital to study the algorithm of data association in order to ensure the accuracy of SLAM state estimation [8], [9], [10], [12]–[14].

In the related research fields of mobile robot SLAM, the data association methods that have been widely applied are independent compatible nearest neighbor (ICNN) algorithm [11], [15] and joint compatible branch and bound (JCBB) algorithm. The advantage of the ICNN algorithm is that the implementation process is relatively easy and it has a strong real-time performance. However, due to not fully consider the correlation of each eigenvalue, its performance will be greatly affected when the environment changes. Compared with the ICNN algorithm, the JCBB algorithm proposed by Castellanos *et al.* [12], [16] improved the correlation accuracy. However, it requires a large amount of computation in the process of algorithm implementation, and its

real-time performance is poor. Researchers continually optimize the data association algorithm, and then put forward the probability data association algorithm, joint maximum likelihood association algorithm and multi-hypothesis tracking algorithm [12], [16]. These algorithms improve the computational efficiency of mobile robot SLAM data association algorithm and solve the data association problem of multi-target tracking [13], [14], [17], [18].

Singer and Stein [15], [19] proposed multiple hypotheses JCBB algorithm to solve the ambiguous data association problem and improved map accuracy. Neira and Tardos [16], [20] adopted mutual exclusion criteria and optimal criteria to improve the accuracy of association. Blackman [17], [21] introduce k-means clustering algorithm to SLAM data association algorithm (KJCBB) for further optimization, which is based on k-means clustering algorithm to cluster the observed values. The number of groups is determined according to the region of the robot. So, it is likely that adjacent landmarks that should be in the same observation group will be misclassified into different observation groups, eventually led to the error association results. Dallil *et al.* [18], [22] proposed a fast JCBB for mobile robot SLAM based on density-based spatial clustering of application with noise. As a result, the association accuracy was improved and the running time of the algorithm was significantly reduced. Wang and Englot [19], [23] proposed a Gaussian mixture model based JCBB data association algorithm (GEMJCBB) grouping the measurement to reduce the computation complexity.

The above studies have contributed to the optimization of SLAM data association algorithm and provided theoretical basis. A new optimized data association algorithm based on Gaussian mixture model [20], [21], [24], [25] is proposed to solve the problem of false association, complicated calculation problem in the current SLAM data association algorithm. Firstly, the local association strategy is adopted to limit data association in local region, so as to reduce the number of features involved in data association at the current moment. Secondly, the Gaussian mixture clustering algorithm is used in local areas to group the observed values at the current moment, so as to get several groups that have little correlation with each other. Finally, JCBB data association algorithm is used in each group for data association, and the optimal solution is obtained according to mutual exclusion criteria and optimal criteria. The algorithm can group the observed data using Gaussian mixture clustering algorithm. Its advantage lies in: 1. Reduce the number of observations and map features participating in association at the same time, thus greatly reducing the computational complexity of JCBB algorithm; 2. The Gaussian mixture model is used to make a reasonable grouping for the local association region, which does not need to be obtained empirically according to the environment, thus improving the accuracy of the algorithm. In section 2, the SLAM data association problem is introduced. In section 3 we introduce the proposed SLAM data association algorithm. Experiment results and analysis are

given in section 4. In addition, the conclusions and the future work are drawn in the last section.

## II. SLAM DATA ASSOCIATION MODEL

### A. SLAM DATA ASSOCIATION

The mobile robot SLAM technology is not only related to the status of the robot itself, but also related to the information of the external environment. The problem of SLAM data association involves three kinds of correspondence. The relationship between the observation obtained by its sensors while the robot is in different time or different region; The relationship between observations and existing environmental map features; The relationship between existing environment map features. Through comparative analysis of the above three relationships, it is determined whether the observations and eigenvalues come from the same entity in the environment region. The above process can be regarded as applying the observation-feature matching character to search the environment state space. There are three possibilities for a robot to get a new observation; One is the built environment characteristics information; Second, the new environmental characteristics; The third is the virtual set, that is, the observed value is not a reflection of real physical road signs, but caused by sensor noise or specular reflection.

Suppose there are  $n$  map features  $F = \{F_1, F_2, \dots, F_n\}$  in the environment area of the robot. The laser sensor measure  $m$  observation  $Z = \{Z_1, Z_2, \dots, Z_m\}$ . The hypothesis needs to be established by applying SLAM data association technology  $H_m = \{j_1, j_2, \dots, j_m\}$ , Matching each observation  $Z_i$  with a feature  $F_{j_i}$ , When the sensor measurements  $Z_i$  do not match any of the features in the map,  $j_i = 0$ . Where, the measured value  $Z_i$  and the corresponding feature  $F_{j_i}$  are related by the measurement function  $f_{j_i}(x, y) = 0$ , indicating that the relative position of the measured value and the corresponding feature must be 0.

### B. JOINT COMPATIBLE BRANCH AND BOUND ALGORITHM

The JCBB algorithm is one of the independent matching association algorithms based on single observation in the research field of mobile robot SLAM. During the implementation of the algorithm, the joint compatibility test method is used to combine the observed features and map features acquired by the mobile robot. While the branch and bound method is used to search the associated solution space.

Under the associated hypothesis set  $H_m = \{j_1, j_2, \dots, j_m\}$ , the joint observation equation of map features is expressed as follows:

$$\hat{Z}_{H_m} = h_{H_m}(\hat{X}_{k,k-1}) + n_{H_m} = \begin{bmatrix} h_{j_1}(\hat{X}_{k,k-1}) \\ \dots \\ h_{j_m}(\hat{X}_{k,k-1}) \end{bmatrix}. \quad (1)$$

Joint new information is expressed by the following formula:

$$S_{H_m} = H_{H_m} P_{k,k-1} H_{H_m}^T + R_{H_m}. \quad (2)$$

Then the covariance of joint new interest can be expressed as:

$$v_{H_m} = Z_{H_m} - h_{H_m}(\hat{X}_{k,k-1}). \quad (3)$$

In (2),  $Z_{H_m} = [Z_{k,1}, \dots, Z_{k,m}]$ ,  $H_{H_m} = \partial h_{H_m} / \partial X|_{(\hat{X}_{k,k-1})}$ .

Then the joint compatible test criterion is as follows:

$$D_{H_m}^2 = v_{H_m}^T s_{H_m}^{-1} v_{H_m}. \quad (4)$$

If the above equation is true, all observed features and map features are considered joint and compatible.

The main purpose of applying branch and bound criterion in the data association problem of mobile robot SLAM is to traverse the solution space and obtain the best solution vector. The joint facultative condition in (4) is used as the criterion for traversing the branch of the correlation interpretation tree. The search order is determined according to its mahalanobis distance, and the whole associated solution space is divided into several subsets. Among the subsets, the monotonic non-subtraction rule of the number of pairs is used as the condition of delimitation. Finally, the largest association hypothesis in the number of pairs is selected as the optimal association solution.

### III. OPTIMIZED DATA ASSOCIATION BASED ON GAUSSIAN MIXTURE CLUSTERING

The observed value solution space in the traditional JCBB SLAM data association algorithm is described according to the interpretation tree model. For obtaining the best data association results, the incremental computational search combining branch and bound with compatibility is adopted. This process takes all sensors into account to obtain the relationship between the observed value. Therefore, the accuracy of data association is higher, and robustness is stronger. But JCBB algorithm associates all of the observed value in the current moment with the environmental characteristics environmental characteristics that already exist on the map. In large dense environments, the number of features increases rapidly over time, resulting in high computational complexity of the algorithm. The algorithm can be improved by reducing the number of observations and map features participating in association at the same time. Firstly, the local association region is obtained to reduce the number of environmental characteristics participating in association at the same time. In addition, the Gaussian mixture clustering algorithm is adopted grouping the observed values to reduce the number of observed values. The specific implementation is as follows:

#### A. LOCAL ASSOCIATION STRATEGY

In the actual situation, the observation range of the robot is limited. It is not necessary for observations to do data association every known characteristics. So, the observed feature that is far away from the robot can be ignored by setting an association threshold in advance to set the features falling within the association threshold as the objects that the target may associate with. The association threshold is set to

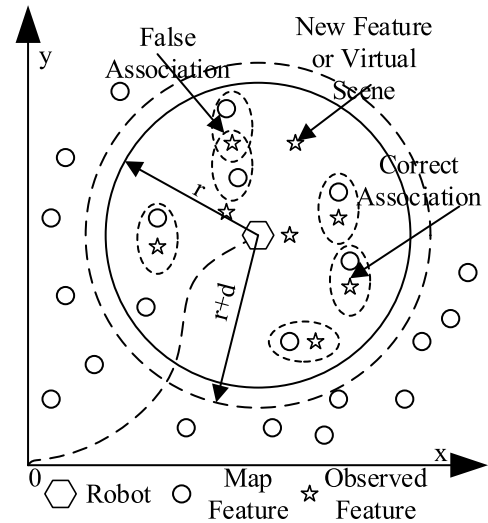


FIGURE 1. Local Data association diagram.

$r + d$ , in which  $r$  is the effective scanning distance of the laser sensor,  $d$  is the distance for compensation. The introduction of compensation distance enables the local map to contain the environmental map features that match the observed values as comprehensively as possible. Local associated areas represent as follows:

$$\sqrt{(x_i - x_r)^2 + (y_i - y_r)^2} < r + d. \quad (5)$$

where:  $(x_f, -y_f)$  represents the position of the feature point, and  $(x_r, -y_r)$  represents the position coordinates of the robot. As shown in Fig. 1, the associated local area is a dotted circle with  $r + d$  as the radius and the robot as the center. Dots represent existing features in the map, and asterisks represent newly observed observations by the sensor. The preprocessing procedure is used to obtain the local association region, so that the number of environmental map features involved in mobile robot SLAM data association can be effectively reduced at a single moment.

#### B. GAUSSIAN MIXTURE CLUSTERING

The observed value generally presents the obvious distribution while mobile robot travel in the environment. In article [21], the k-means clustering algorithm are applied to group the observed data. It is easy to lead to local optimization while the overall observed data is large. And meanwhile, when dealing with practical problems, the algorithm is sensitive to noise and outliers. K-means is only applicable to numerical type data. However, non-convex data and irregular shape of the cluster cannot be solved effectively. The article [22] then use density-based method system to solve this problem. This method can not only find any shape of clustering, but also process the data with noise more accurately. The results obtained by density-based clustering method are related to the fixed parameters used to identify clustering. The algorithm will still be affected by the sparsity even with the same standard. In other words, it will be divided into

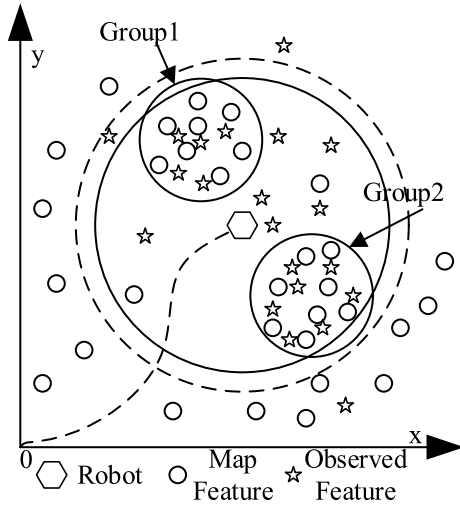


FIGURE 2. The clustering grouping diagram of data association.

multiple classes if the data is relatively sparse and merged to one if relatively dense. In recent years, we have witnessed several researches on clustering algorithm related to GMM, such as [22]–[31]. The probability density function in the Gaussian mixture model (GMM) plays a very important role in simplifying the processing steps of data and allocating accurate processing results to each Gaussian mixture. In the process of parameter estimation, GMM applies the maximum expectation (EM) algorithm, which improves the operation efficiency of data analysis [32], [33]. Therefore, we adopt Gaussian mixture clustering algorithm to group the observed values. The clustering grouping diagram of data association is shown in Fig. 2.

Firstly, the observed eigenvalue set  $Z = \{Z_1, Z_2, \dots, Z_m\}$  and Gaussian mixture number  $k$  were taken as the input, and then the Gaussian mixture distribution model was initialized. The selection of  $k$  value in the Gaussian mixture model is determined according to the environment. According to the range of observed values scanned by the laser, the number of groups is generally 3 to 5. The Gaussian mixture distribution (GMM) considers that data is generated from several single Gaussian distribution models (GSM), sigma is the variance of the model. The probability density function is

$$\Pr(z_i) = \sum_{i=1}^k \pi_k N(z_i; \mu_i, \sigma_i) \dots \quad (6)$$

where  $\pi_k$  is the weight factor, represents the probability of choosing the  $i$ th mixed ingredient, and  $\sum_{i=1}^k \pi_i = 1$ . Single Gaussian distribution  $N(z; \mu, \sigma)$  represents a cluster grouping in the Gaussian mixture distribution, as follows:

$$N(z; \mu, \sigma) = \frac{1}{\sqrt{2\pi} |\sigma|} \cdot e^{\left[-\frac{1}{2}(z-\mu)^T \sigma^{-1}(z-\mu)\right]}. \quad (7)$$

The Gaussian mixture component of the generated sample  $z_j (j = 1, \dots, m)$  is expressed as a random variable  $q_j$ , and the prior probability  $P(q_j = i)$  of  $q_j$  corresponds to

$z_j (j = 1, \dots, m)$ . According to Bayes theorem, the posterior probability density function of sample  $z_j (j = 1, \dots, m)$  generated according to the  $i$ th GMM is now expressed as  $z_j (j = 1, \dots, m)$ :

$$\gamma_{ji} = \frac{\pi_i \cdot p(z_j | \mu_i, \sigma_i)}{\sum_{l=1}^k \pi_l \cdot p(z_j | \mu_l, \sigma_l)}. \quad (8)$$

Gaussian mixture clustering divides the observation feature set  $Z$  into  $k$  components, denoted by  $Group = \{Group_1, \dots, Group_k\}$ , and the component mark of each observation sample is denoted by:

$$\lambda_j = \arg \max_{i \in \{1, \dots, k\}} \gamma_{ji}. \quad (9)$$

Gaussian mixture model is a clustering algorithm with  $k$  clustering centers. The calculation performance of Gaussian function is relatively good. If the classification of the samples in the system is unknown, it is possible to calculate  $(\pi, \mu, \sigma)$  given only the sample points. For a given observation data set  $Z$ , the model parameters can be solved by maximum likelihood estimation.

The number of data information points observed by the laser sensor is  $m$ , and it obeys a certain distribution  $\Pr(z_i; \lambda)$ . The goal of the algorithm is to get the parameter  $\lambda$  so that the probability  $\prod_{i=1}^m \Pr(z_i; \lambda)$  of the generated data point is maximize, where  $\prod_{i=1}^m \Pr(z_i; \lambda)$  is called the likelihood function. In general, the probability value of a single data point is relatively small, and the product of the probability values of  $m$  data points is smaller. So floating point underflow is very easy to occur. Therefore, the logarithm of likelihood function is normally taken as  $\sum_{i=1}^m \ln \Pr(z_i; \lambda)$  known as log-likelihood function. The log-likelihood function of GMM is as follows:

$$LL(Z) = \sum_{j=1}^m \ln \left( \sum_{i=1}^k \pi_i \cdot p(z_j | \mu_i, \sigma_i) \right). \quad (10)$$

Therefore, the maximum expected value algorithm can be applied to solve the iterative optimization of the observation data set model, and then the classification of observation data samples can be distinguished according to the model that has completed training. The specific steps are expressed as follows: Firstly, select one randomly from the groups of  $k$  clusters; Then, the observed data samples were substituted into the cluster selected in the previous step, and whether the data samples belonged to this category was determined. If not, the cluster group was re-selected.

Expectation step: assuming that model parameters are known, calculate the Expectation of implicit variables taking  $q_1, q_2, \dots$  respectively, that is, the probability of  $Q$  Taking  $q_1, q_2, \dots$  respectively. This step in the Gaussian mixture model is to calculate the probability  $\gamma_{ji}$  of the observed sample data points generated by each cluster group.

Maximization step: maximum likelihood method is considered to calculate model parameters. Take the  $\gamma_{ji}$  obtained

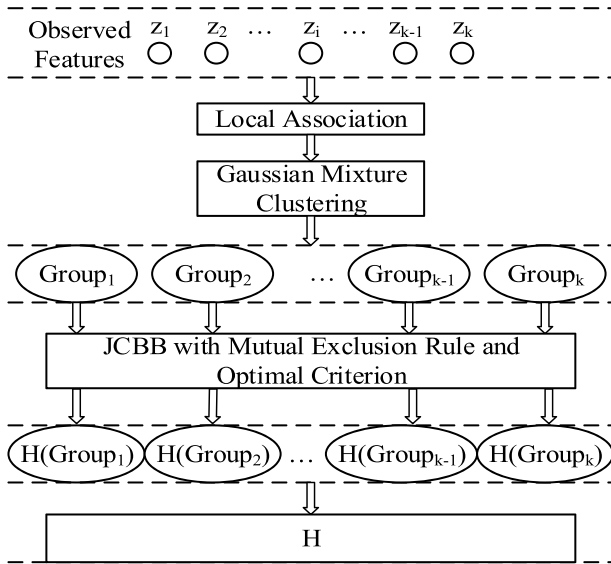


FIGURE 3. GEOJCBB algorithm flow chart.

by the Expectation operation step as the probability that the observed value sample data point  $Z_i$  is generated by grouping the  $k$ th cluster. If  $LL(Z)$  is maximized by  $(\pi_i, \mu_i, \sigma_i)$ , then we can let  $\partial LL(Z) / \partial \mu_i = 0$  to get

$$\mu_i = \frac{1}{\sum_{j=1}^m \gamma_{ji}} \cdot \sum_{j=1}^m \gamma_{ji} z_j. \quad (11)$$

The sample weight is the posterior probability of each sample belonging to the component, and let  $\partial LL(Z) / \partial \sigma_i = 0$  to get

$$\sigma_i = \frac{1}{\sum_{j=1}^m \gamma_{ji}} \cdot \sum_{j=1}^m \gamma_{ji} (z_j - \mu_i) (z_j - \mu_i)^T. \quad (12)$$

$\pi_i > 0$  and,  $\sum_{i=1}^k \pi_i = 1$ , maximize  $LL(Z)$  and get

$$\pi_i = \frac{1}{m} \cdot \sum_{j=1}^m \gamma_{ji}. \quad (13)$$

According to (11) to (13) to update the model parameters, and according to (9) to determine the component  $z_j$ , divide  $z_j$  into the corresponding component, namely  $Group_{\lambda_j} = Group_{\lambda_j} \cup \{z_j\}$ , and finally get the division result of component  $Group = \{Group_1, \dots, Group_k\}$ .

Every step of the mobile robot motion, through the GMM algorithm about all of the observed value is divided into several small correlation grouping. Then for each grouping JCBB data correlation method is used to get local association results. Each Group represented as explained in the tree layer. And Select the best pairings from the explanation tree as the final result. In this paper, the abbreviation of the proposed algorithm is GEOJCBB. The algorithm description is shown in Fig. 3.

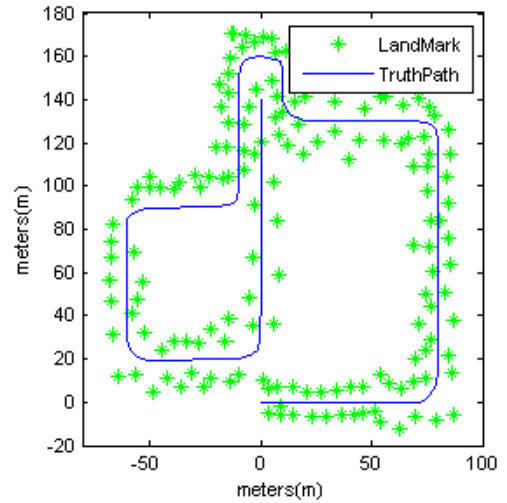


FIGURE 4. Simulation environment.

### C. OPTIMIZED METHOD

In practice, JCBB data association have false association problem. One is that different observed features match the same features on the map. The other is that JCBB accepts an associative solution that is not optimal when there are multiple associative solutions with the maximum matched logarithm. In order to address the problem, we introduce mutual exclusion rule and optimal criterion.

Mutual exclusion rule: a map feature is allowed to be associated with only one observed value. This means that after a map feature is matched, another observation will be rejected if it is matched with the map feature again, and the observation can only look for the unmatched map feature to be correlated or treated as a new feature.

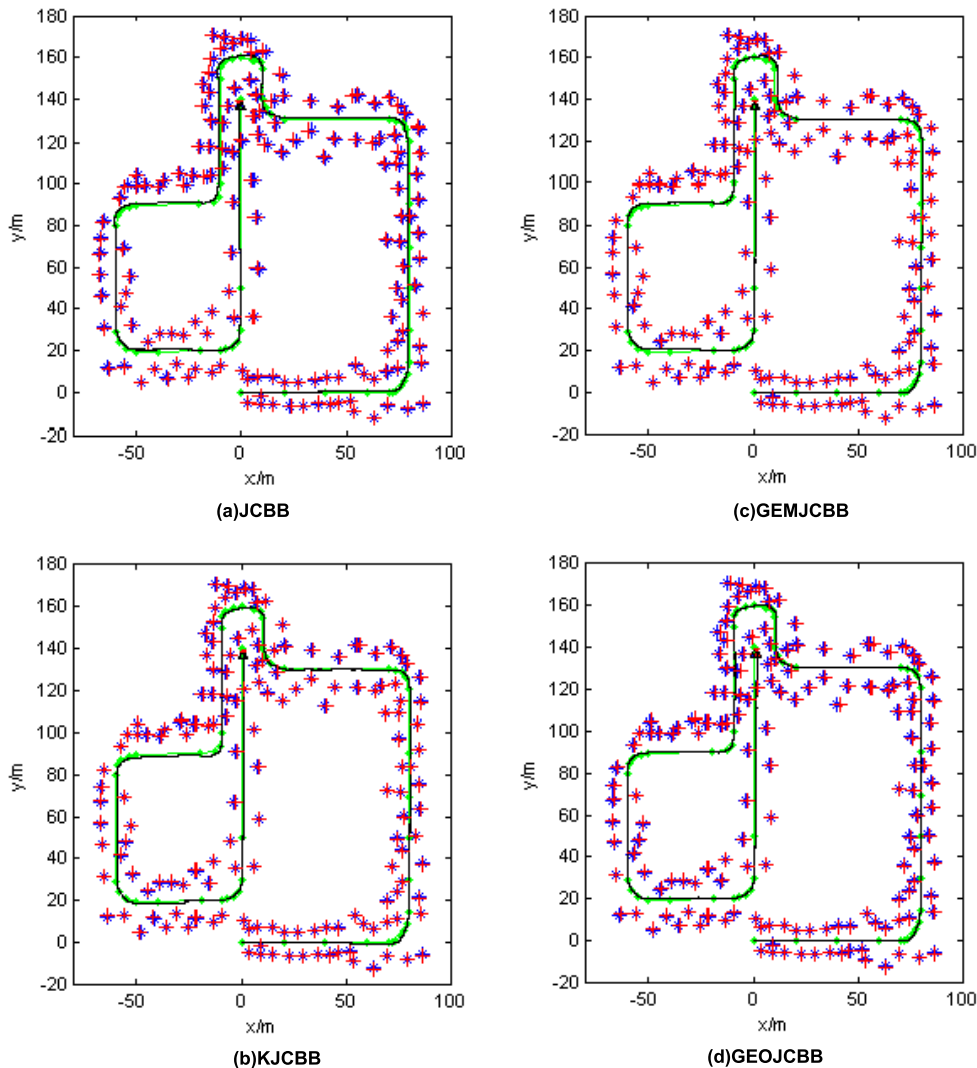
Optimal criterion: if the pairing number of multiple associative solutions is equal to the maximum pairing number, the associative solution with the minimum joint mahalanobis distance is selected as the final associative result.

$$H = H^k |_{(D^2_{(H^k)} \leq D^2_{(H^i)}), \quad i = 1, \dots, n. \quad (14)$$

where,  $n, H^i, H^k$  represents the maximum pairing number, represents the  $i$ th associative solution and the optimal associative solution.

### IV. EXPERIMENTAL RESULTS AND ANALYSIS

The simulation environment shown in Fig. 4 was designed on the simulation experiment platform developed by Neira *et al.* [34] to verify the improvement in association performance of the GEOJCBB data association algorithm proposed in this paper. The "\*" in the figure represents the existing features in the environmental space, and the solid line represents the theoretical planning path of the robot. In order to verify the high efficiency and accuracy of the algorithm in the association results in this paper, the positioning accuracy, association time complexity and correlation performance were respectively compared and analyzed with JCBB data



**FIGURE 5.** Path fitting of four data association methods in simulation environment.

association algorithm, KJCBB data association algorithm and GEMJCBB data association algorithm.

### A. PATH FITTING RESULT

The data association and path fitting results of JCBB data association algorithm, KJCBB data association algorithm, GEMJCBB data association algorithm, and GEOJCBB data association algorithm are shown in Fig. 5. In Fig. 5, a, b, c, and d correspond to JCBB data association algorithm, KJCBB data association algorithm, GEMJCBB data association algorithm, and GEOJCBB data association algorithm respectively. The solid green line is the actual path of the robot. The blue “\*” indicates the actual features in the environment. The red “+” is the predictive feature; and the black line is the estimated path.

As is shown in Fig. 5, the estimated path fits the true path well in the four algorithms and have no obvious difference visually.

The error curves of pose estimation in Fig. 6 and Fig. 7 describe the X-axis and Y-axis errors of the actual path and the estimated path in simulation environment. The mean estimation errors of SLAM in X-axis direction based on JCBB, KJCBB, GEMJCBB, and GEOJCBB algorithms were respectively 0.5184, 0.4352, 0.3895, and 0.3029. The mean estimation errors in the Y-axis direction were 0.6306, 0.5450, 0.3951, and 0.2389. Simulation results show that the GEOJCBB based SLAM data association algorithm proposed in this paper provides more reliable association results, which improves the estimation accuracy of the pose of SLAM compared with the other three algorithms.

### B. ASSOCIATION EFFICIENCY

In the simulation environment, the robot moves counterclockwise and uniformly from the initial state. Twenty Monte Carlo simulation experiments were conducted for SLAM based on

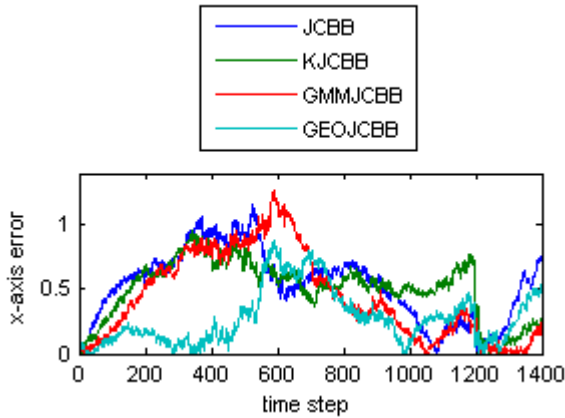


FIGURE 6. Path fitting error of x-axis.

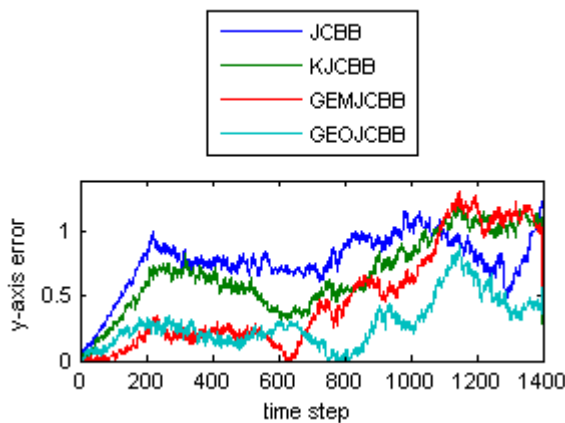


FIGURE 7. Path fitting error of y-axis.

TABLE 1. Comparison of algorithm association time in simulation environment.

Environment	Algorithm	Association Time
Environment	JCBB	210.7598
	KJCBB	155.0357
	GEMJCBB	144.8525
	GEOJCBB	135.0218

four association algorithms. Table. 1 shows the average association time of four association algorithms in 20 simulation experiments.

According to Table. 1, the average association time of the GEOJCBB SLAM data association algorithm proposed in this paper in simulation environment is 135.0218seconds. The average association time of the algorithm in The analysis of the simulation results shows that the average association time of the algorithm proposed in this paper are both less than the other three algorithms. Specifically, KJCBB, GEMJCBB, and GEOJCBB all grouped observations, reducing the dimension of observation measurement in the joint compatible calculation, thus reducing the complexity of calculation. Meanwhile, GEOJCBB adopts Gaussian mixture

model to cluster and group the observed values. The probability density function in GMM can accurately distribute the observed sample data to each mixed component and simplify the data processing steps. At the same time, GMM referred to the idea of EM algorithm when evaluating parameters, which significantly improved the data analysis speed of the algorithm. In addition, the GEOJCBB algorithm proposed in this paper delimits the local association area in the data pre-processing operation, reducing the number of environmental features participating in the data association at the same time, so as to further improve the efficiency of the algorithm.

C. ASSOCIATION PERFORMANCE

The four indexes to evaluate the performance of data association are: the accuracy of association (TP rate, TPR), the accuracy of new road signs added to the map (TN rate, TNR), the error rate of association (FP rate, FPR) and the missing rate of association (FN rate, FNR). Where, TP, TN, FP and FN are defined as follows: TP: true positive, means the correct measurement-characteristic matching logarithm detected. TN: true negative, which represents the unmatched logarithm of the correct rejection, namely the number of new environment features detected. FP: false positive, indicating the measurement of detected errors-logarithm of feature matching. FN: false negative means the correct measurement-characteristic matching logarithm that was not detected.

The number of observations at the current moment is Total:

$$Total = TP + TN + FP + FN. \tag{15}$$

Precision is the percentage of the samples with positive prediction. It is calculated as follows:

$$Precision = \frac{TP}{TP + FP}. \tag{16}$$

Recall, True Positive Rate (TPR), represent the Positive sample Rate in the judgment pair samples, that is, the proportion of the detected correct observation-feature matching logarithm in the total observation-feature matching logarithm, and the calculation formula is as follows:

$$Recall = \frac{TP}{TP + FN}. \tag{17}$$

In order to make a more comprehensive and scientific evaluation algorithm, the most common method is F-measure. F-measure is the weighted harmonic average of Precision and Recall:

$$F = \frac{(\alpha^2 + 1) * Precision * Recall}{Precision + Recall}. \tag{18}$$

In (19),  $\alpha$  equals 1. Therefore, the F-measure in this paper is:

$$F = \frac{2 * Precision * Recall}{Precision + Recall}. \tag{19}$$

Fig. 8 shows the data association performance of JCBB, KJCBB, GEMJCBB, and GEOJCBB algorithms in simulation environment.

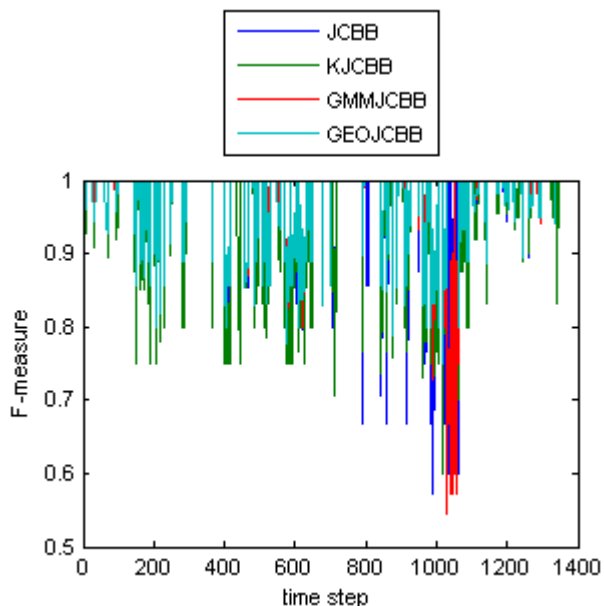


FIGURE 8. Data association performance.

TABLE 2. Comparison of algorithm F-measure in simulation environment.

Environment	Algorithm	F-measure
Environment	JCBB	0.9619
	KJCBB	0.9651
	GEMJCBB	0.9714
	GEOJCBB	0.9781

Table. 2 shows the average F-measure of four association algorithms. As shown in Table. 2, the F-measure of the proposed algorithm in the simulation environment is improved compared with the other three algorithms.

V. CONCLUSION

Data association is the foundation of state estimation in mobile robot SLAM JCBB algorithm is currently a common SLAM data association algorithm that can be obtain reliable results. But in a wide range of dense environment, the number of map environment features and observed values participating in the association increases rapidly with time, leading to the increase of computation amount of the JCBB algorithm. In practical application, a good data association algorithm should not only have high accuracy, but also the characteristics of superior real-time performance and low computational complexity. This paper presents a SLAM data association algorithm based on Gaussian mixture clustering. The algorithm can be improved by reducing the number of observations and map features participating in the association at the same time. Firstly, the local association strategy is adopted to limit data association in local region, so as to reduce the number of features involved in data association at the current moment. Secondly, the Gaussian mixture clustering algorithm is used in local areas to group the observed values at the current moment, so as to get several groups that

have little correlation with each other. Finally, JCBB data association algorithm is used in each group for data association, and the optimal solution is obtained according to mutual exclusion criteria and optimal criteria. Experimental results show that the proposed JCBB algorithm based on Gaussian mixture clustering optimization can obtain accurate association results, reduce computational complexity, improve algorithm efficiency, and provide a reliable guarantee for localization of mobile robot SLAM. However, the experiment was realized in simulation environment. In the future, we will implement the experiment in the real word.

ACKNOWLEDGMENT

The authors thank P. Li, X. Chen, and C. Tan for their valuable discussion and support. And the authors thank anonymous reviewers for giving them help to improve the quality of the paper greatly.

REFERENCES

- [1] J. J. Leonard and H. F. Durrant-Whyte, "Simultaneous map building and localization for an autonomous mobile robot," in *Proc. IEEE/RSJ Int. Workshop Intell. Robot. Syst. (IROS)*, Osaka, Japan, Nov. 1991, pp. 1442–1447.
- [2] M. W. M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Trans. Robot. Autom.*, vol. 17, no. 3, pp. 229–241, Jun. 2001, doi: 10.1109/70.938381.
- [3] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–110, Jun. 2006, doi: 10.1109/mra.2006.1638022.
- [4] S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, "Simultaneous localization and mapping with sparse extended information filters," *Int. J. Robot. Res.*, vol. 23, nos. 7–8, pp. 693–716, Aug. 2004, doi: 10.1177/0278364904045479.
- [5] 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.
- [6] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Proc. Robot. Sci. Syst.*, Berkeley, CA, USA, 2014.
- [7] J. Engel J, T. Schöps, Cremers D. "LSD-SLAM: Large-scale direct monocular SLAM," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 834–849.
- [8] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D Cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017, doi: 10.1109/tro.2017.2705103.
- [9] M. Labb e and F. Michaud, "RTAB-Map as an open-source Lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *J. Field Robot.*, vol. 36, no. 2, pp. 416–446, 2019.
- [10] C. Yu, Z. Liu, X.-J. Liu, F. Xie, Y. Yang, Q. Wei, and Q. Fei, "DS-SLAM: A semantic visual SLAM towards dynamic environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, Madrid, Spain, Oct. 2018, pp. 1168–1174.
- [11] B. Bescos, J. M. F acil, J. Civera, and J. Neira, "DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4076–4083, Oct. 2018.
- [12] J. A. Castellanos, J. D. Tardos, and G. Schmidt, "Building a global map of the environment of a mobile robot: The importance of correlations," in *Proc. Int. Conf. Robot. Autom.*, Albuquerque, NM, USA, Apr. 1997, pp. 1053–1059.
- [13] P. S. Segundo and D. Rodriguez-Losada, "Robust global feature based data association with a sparse bit optimized maximum clique algorithm," *IEEE Trans. Robot.*, vol. 29, no. 5, pp. 1332–1339, Oct. 2013, doi: 10.1109/tro.2013.2264869.
- [14] S. Ivaldi, S. Lefort, J. Peters, M. Chetouani, J. Provasi, and E. Zibetti, "Towards engagement models that consider individual factors in HRI: On the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task," *Int. J. Soc. Robot.*, vol. 9, no. 1, pp. 63–86, Jun. 2016, doi: 10.1007/s12369-016-0357-8.



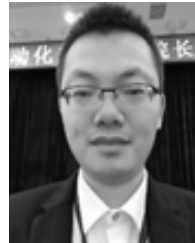
- [15] R. A. Singer and J. J. Stein, "An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance systems," in *Proc. IEEE Conf. Decis. Control*, Miami Beach, FL, USA, Dec. 1971, pp. 171–175.
- [16] J. Neira and J. D. Tardos, "Data association in stochastic mapping using the joint compatibility test," *IEEE Trans. Robot. Autom.*, vol. 17, no. 6, pp. 890–897, Dec. 2001, doi: [10.1109/70.976019](https://doi.org/10.1109/70.976019).
- [17] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 5–18, Jan. 2004, doi: [10.1109/maes.2004.1263228](https://doi.org/10.1109/maes.2004.1263228).
- [18] A. Dallil, M. Oussalah, and A. Ouldali, "Sensor fusion and target tracking using evidential data association," *IEEE Sensors J.*, vol. 13, no. 1, pp. 285–293, Jan. 2013, doi: [10.1109/jsen.2012.2213892](https://doi.org/10.1109/jsen.2012.2213892).
- [19] J. Wang and B. Englot, "Robust exploration with multiple hypothesis data association," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, Madrid, Spain, Oct. 2018, pp. 3537–3544.
- [20] W. Zhou and C. X. Zhao, "Optimized data association solution for SLAM," *Robot.*, vol. 31, no. 3, pp. 217–223, Jan. 2009.
- [21] C. Yao, *The Research of Mobile Robot Localization Method Based on Bayesian Theory*. Chongqing, China: Chongqing Univ., 2015.
- [22] D. Liu, J. Duan, and C. Wang, "A fast joint compatibility data association algorithm for SLAM based on clustering," *Robot.*, vol. 40, no. 2, pp. 158–177, Mar. 2018, doi: [10.13973/j.cnki.robot.170400](https://doi.org/10.13973/j.cnki.robot.170400).
- [23] X. Ruan, J. Zhang, and Z. Xiaoqing, "Simultaneous localization and mapping data association based on maximum expectation clustering for Gaussian mixture model," *Control Theory Appl.*, pp. 1–11, Jul. 2019.
- [24] T. Nagai, Y. Mochizuki, Y. Joti, F. Tama, and O. Miyashita, "Gaussian mixture model for coarse-grained modeling from XFEL," *Opt. Express*, vol. 26, no. 20, pp. 26734–26749, Oct. 2018, doi: [10.1364/oe.26.026734](https://doi.org/10.1364/oe.26.026734).
- [25] M. C. Madhavi and H. A. Patil, "Design of mixture of GMMs for Query-by-example spoken term detection," *Comput. Speech Lang.*, vol. 52, pp. 41–55, Nov. 2018, doi: [10.1016/j.csl.2018.04.006](https://doi.org/10.1016/j.csl.2018.04.006).
- [26] Y. Yu, M. Li, and Y. Fu, "Forest type identification by random forest classification combined with SPOT and multitemporal SAR data," *J. For. Res.*, vol. 29, pp. 1407–1414, Sep. 2017, doi: [10.1007/s11676-017-0530-4](https://doi.org/10.1007/s11676-017-0530-4).
- [27] A. Nazari, A. Dehghan, S. Nejatian, V. Rezaie, and H. Parvin, "A comprehensive study of clustering ensemble weighting based on cluster quality and diversity," *Pattern Anal. Appl.*, vol. 22, no. 1, pp. 133–145, Feb. 2019, doi: [10.1007/s10044-017-0676-x](https://doi.org/10.1007/s10044-017-0676-x).
- [28] S. Mingjie and C. Dongmei, "An improved knowledge-informed NSGA-II for multi-objective land allocation (MOLA)," *Geo-Spatial Inf. Sci.*, vol. 21, no. 4, pp. 273–287, Oct. 2018, doi: [10.1080/10095020.2018.1489576](https://doi.org/10.1080/10095020.2018.1489576).
- [29] A. Sadr-Olah, N. Samad, P. Hamid, R. Vahideh, and B. Karamolah, "Clustering ensemble selection considering quality and diversity," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 1311–1340, Aug. 2018, doi: [10.1007/s10462-018-9642-2](https://doi.org/10.1007/s10462-018-9642-2).
- [30] G. Bordogna, S. Capelli, D. E. Ciriello, and G. Psaila, "A cross-analysis framework for multi-source volunteered, crowdsourced, and authoritative geographic information: The case study of volunteered personal traces analysis against transport network data," *Geo-Spatial Inf. Sci.*, vol. 21, no. 3, pp. 257–271, Oct. 2017, doi: [10.1080/10095020.2017.1374703](https://doi.org/10.1080/10095020.2017.1374703).
- [31] M. J. Hosseinpoor, H. Parvin, S. Nejatian, and V. Rezaie, "Gene regulatory elements extraction in breast cancer by Hi-C data using a meta-heuristic method," *Russ. J. Genet.*, vol. 55, no. 9, pp. 1152–1164, Nov. 2019, doi: [10.1134/s1022795419090072](https://doi.org/10.1134/s1022795419090072).
- [32] H. Parvin and B. Minaei-Bidgoli, "A clustering ensemble framework based on elite selection of weighted clusters," *Adv. Data Anal. Classification*, vol. 7, no. 2, pp. 181–208, Apr. 2013, doi: [10.1007/s11634-013-0130-x](https://doi.org/10.1007/s11634-013-0130-x).
- [33] H. Parvin and B. Minaei-Bidgoli, "A clustering ensemble framework based on selection of fuzzy weighted clusters in a locally adaptive clustering algorithm," *Pattern Anal. Appl.*, vol. 18, no. 1, pp. 87–112, Jan. 2014, doi: [10.1007/s10044-013-0364-4](https://doi.org/10.1007/s10044-013-0364-4).
- [34] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, "Consistency of the EKF-SLAM algorithm," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Beijing, China, Oct. 2006, pp. 3562–3568.



**XIAOGANG RUAN** was born in Sichun, China. He received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 1992. He is currently a Professor with the Beijing University of Technology and also the Director of the Institute of Artificial Intelligent and Robots (IAIR). His research interests include automatic control, artificial intelligence, and intelligent robot.



**DINGQI REN** was born in Hebei, China. He is currently pursuing the master's degree with the Beijing University of Technology. His research interests include robotics and artificial intelligence, and the main research work now is robot navigation problem.



**XIAOQING ZHU** received the Ph.D. degree in control science and engineering from the Beijing University of Technology, in 2015. He is currently the Deputy Dean of College of Artificial Intelligence and automation with the Faculty of Information Technology, Beijing University of Technology, and he is also a Guest Professor with the College of Electronic and Information, Nanchang Institute of Technology. His research interests include robotics and machine learning.



**SHAODA LIU** was born in Hebei, China. He is currently pursuing the master's degree with the Beijing University of Technology. His research interests include robotics and artificial intelligence, and the main research work now is robot navigation problem.

...